PROCESSING HELIUM-SPEECH: PROCEEDINGS OF A NAVY-SPONSORED WORKSHOP AUGUST 1971

Edited by
Russell L. Sergeant, Ph.D.
and
LT Thomas Murry, MSC, USNR

NAVAL SUBMARINE MEDICAL RESEARCH LABORATORY
NAVAL SUBMARINE MEDICAL CENTER REPORT NO. 708

Bureau of Medicine and Surgery, Navy Department
Research Work Unit M4306.03-2020DAC5.10

and

Office of Naval Research
Research Project No. RR 042-09-02
NR 197-0003

Transmitted by:

J. Donald Harris

J. Donald Harris, Ph.D.
Head, Auditory Research Branch

Reviewed and Approved by:

Charles F. Gell

Charles F. Gell, M.D., D.Sc. (Med)
Scientific Director
NavSubMedRschLab

Approved and Released by:

R. L. Sphar

R. L. Sphar, CDR, MC, USN
Officer in Charge, Acting
NavSubMedRschLab

SUMMARY PAGE

THE PROBLEM

To evaluate the present status of equipment for processing helium-speech and to assess expected developments toward reliable communication by voice within hyperbaric helium-oxygen environments.

FINDINGS

A workshop was conducted on helium-speech processing, attended by foreign scientists, U. S. Navy scientists, operational personnel, Naval and independent contractors, and speech scientists in the academic world, who have all been active in underwater communications. Ten papers were presented, a forum and discussion were held, and a summary and comments were presented. It was concluded that correction of hyperbaric helium speech finally can be accomplished. It was concluded that a system that is small, inexpensive and reliable must be designed and incorporated into diving operations.

APPLICATION

Information contained in this report is useful to the design of systems intended to improve the voice communicability of divers who operate within hyperbaric helium atmospheres.

ADMINISTRATIVE INFORMATION

This investigation was conducted as part of Bureau of Medicine and Surgery Research Work Unit M4306.03-2020DAC5 -- Evaluation of Underwater Communication Systems for Navy Divers, and the Office of Naval Research Project No. RR 042-09-02, NR 197-0003 - Conduct of a workshop on helium-speech processing. The present report is No. 10 on this Work Unit. It was approved for publication of 22 May 1972 and designated as Naval Submarine Medical Research Laboratory Report No. 708.

DISCLAIMER STATEMENT

PUBLISHED BY THE NAVAL SUBMARINE MEDICAL RESEARCH LABORATORY

ABSTRACT

This report is a detailed summary of the proceedings of a workshop held during August 1971 on helium-speech processing. The meeting was jointly sponsored by the Office of Naval Research and the Bureau of Medicine and Surgery. It was held at the Naval Submarine Medical Research Laboratory in Groton, Connecticut. Approximately 40 participants were brought together, including foreign scientists, U. S. Navy scientists, operational personnel, Naval and independent contractors, and speech scientists in the academic world, all who have been active in underwater communications. Formal papers were presented and discussed, a forum and discussion period was held, and a summary and comments were presented. Progress and future developments toward reliable speech communication under hyperbaric helium-oxygen conditions were assessed. Concepts of helium-speech processing were advanced from the need of an unscrambler unit to one which includes the understanding and nature of the constraints allied to the unscrambler; that is, talker, listener, face mask and transducer. It was concluded that after a decade of research the ability to correct hyperbaric helium speech finally exists. Now a system that is small, inexpensive, rugged and reliable must be designed and incorporated into diving operations.

TABLE OF CONTENTS

TABLE OF CONTENTS (cont)

PREFACE

Initial recognition of the need for a workshop on processing helium speech evolved from discussions among personnel from the Bureau of Medicine and Surgery, the Office of Naval Research, and the Naval Submarine Medical Research Laboratory. From these discussions, it was concluded that an extensive, yet fragmented, body of information existed in the area of helium speech processing and underwater voice communication. This workshop, therefore, was undertaken to bring together interested scientists and military personnel in order to unify present knowledge and indicates directions for future research and development. Dr. Charles F. Gell, Scientific Director of the Submarine Medical Research Laboratory and Dr. Gilbert C. Tolhurst, former head of the Physiological Psychology Branch at the Office of Naval Research, inspired the initial proposal for the workshop, and thereafter provided guidance through its formative stages to successful completion of the program. Selection of major speakers and participants was made on the basis of their particular active interests in underwater voice communication with special emphasis on how to process helium speech. Thus, the scope of the subject matter, while being somewhat restrictive, nevertheless extends to all phases of the problem of underwater voice communication: history, preliminary approaches to the problems of unscrambling speech, theoretical implications and solutions, phonology, systems design and operational requirements.

The report of the proceedings which follows was based on edited recordings made during the sessions as well as manuscripts submitted by each of the major contributors. The overall character represents a single, current report about processing helium speech and critical references thereto. The report is especially timely in that it occurs at a major turning point in the advancement of state-of-the-art of processing the speech of divers and swimmers. As such, it is indeed a valuable source for scientists, engineers and the operating forces of the U. S. Navy.

J. D. Bloom, CDR, MC, USN
Officer-in-Charge
Naval Submarine Medical Research
Laboratory

PARTICIPANTS

Name	Activity Represented
Mr. J. Abbruzo	Westinghouse Corporation Baltimore, Maryland
Mr. C. Allen	Naval Electronics Laboratory Center San Diego, California
Dr. D. Baker	University of Maryland College Park, Maryland
Dr. B. Bauer	CBS Laboratories Stamford, Connecticut
LT T. Berghage, MC, USN	Navy Experimental Diving Unit Washington, D.C.
CDR J. D. Bloom, MC, USN	Naval Submarine Medical Research Lab. Groton, Connecticut
Dr. R. Carr	Raytheon Company Portsmouth, Rhode Island
Mr. W. Carpenter	Naval Electronics Laboratory Center San Diego, California
LT D. Chandler, MSC, USN	Navy Experimental Diving Unit Washington, D. C.
LT E. Choppell	Chief of Naval Operations Washington, D. C.
Dr. R. Coleman	Bill Wilkerson Hearing and Speech Center Nashville, Tennessee
Dr. D. Darling	Raytheon Company Portsmouth, Rhode Island
LCDR P. Dowland, RN	U.S. Navy Supervisor of Diving Naval Ship Systems Command Washington, D. C.

PARTICIPANTS

Name	Activity Represented
Mr. J. Elkins	Naval Coastal Systems Laboratory Panama City, Florida
Mr. R. Flower	Singer-General Precision, Kearfott Div. Pleasantville, New York
Dr. J. S. Gill	Admiralty Research Laboratory Teddington, England
Mr. J. Helle	Helle Engineering Inc. San Diego, California
Dr. J. D. Harris	Naval Submarine Medical Research Lab. Groton, Connecticut
Mr. J. Hudgens	Westinghouse Corporation Baltimore, Maryland
Mr. J. Jenkins	Naval Coastal Systems Laboratory Panama City, Florida
Dr. J. Lindqvist	KTH Speech Transmission Laboratory Stockholm, Sweden
Mr. J. Malecki	Office of Naval Research Washington, D.C.
Dr. C. Morrow	Advanced Technology Laboratory Houston, Texas
Mr. W. Mullen	Naval Coastal Systems Laboratory Panama City, Florida
Dr. T. Murray	KTH Speech Transmission Laboratory Stockholm, Sweden
LT T. Murry, MSC, USNR	Naval Submarine Medical Research Lab. Groton, Connecticut

PARTICIPANTS

Name	Activity Represented
CDR P. Nelson, MSC, USN	Bureau of Medicine and Surgery Department of the Navy Washington, D.C.
Dr. J. O'Hare	Office of Naval Research Washington, D.C.
LT G. Olson, MSC, USNR	Naval Submarine Medical Research Lab. Groton, Connecticut
Dr. A. Oppenheim	Massachusetts Institute of Technology Cambridge, Massachusetts
LCDR H. Panek, MC, USN	Bureau of Medicine and Surgery Department of the Navy Washington, D.C.
LCDR A. Peck, MC, USN	Naval Submarine Medical Research Lab. Groton, Connecticut
Mr. A. Pesch	Electric Boat Division, General Dynamics Groton, Connecticut
Mr. H. Reuter	Naval Ship Systems Command Washington, D.C.
Dr. R. Robertson	Naval Aerospace Medical Center Pensacola, Florida
Mr. F. Romano	Naval Ship Systems Command Washington, D.C.
Dr. H. Rothman	University of Florida Gainesville, Florida

placed in the mask or hard hat. The frequency output of the resonators suitably shifted downward would then be summed and amplified. The space requirements seemed excessive.

e. Various configurations of vocoder techniques have been advanced by a number of investigators. In theory, a design which would utilize formant vocoder tracking should result in an output that would not only be intelligible but also one that would preserve more of the speaker's individual characteristics, i.e., sound more natural. Until two or three years ago such a system would have to be excessively large, but microminiaturization techniques now include both active and passive filtering. This methodology may be the next logical design step that should be attempted. There probably will be other ideas, circuitry and procedures arising from the contributions of the participants of this workshop.

If there are any schemes, instruments, or developments not mentioned in this brief history, or given too short or cursory description, it was truly unintentional.

3. INTELLIGIBILITY AND PERCEPTUAL ASPECTS OF HELIUM SPEECH

by Robert F. Coleman

In evaluating a communication system, the entire chain between a talker and listener must be included, or erroneous estimates of performance can be obtained. Thus, testing of hardware is fine, but represents the first step in an overall evaluation of the system. In addition, a significant amount of research now available indicates that the speaker, particularly, and the listener to a lesser degree are responsible for a good deal of the poor performance of commercially available communication systems. The initial coupling of the diver/talker to the transducing system, at least in air mixtures using conventional microphones, results in approximately a 20% degradation of the message to be transmitted. Recent developments by Morrow, et al [32,33] appear to indicate that this coupling effect can be compensated for by microphone design or other system changes.

"Intelligibility" refers primarily to a property of speech communication involving meaning, rather than simple recognition of specific speech sounds. Since "meaning" is a wastebasket term for human processing of speech, it is logical that there are many different types of meaning which are not necessarily equivalent. The upshot of this rather academic discussion is that we must be very cautious about assuming that because a communicator achieves a score of 90% on a monosyllabic word test, that it will be "90% intelligible" in all situations. Such is not the case.

A major point I wish to make is that intelligibility scores are relative rather than absolute; the further you extrapolate from the test material used, the greater the likelihood that you will make an error in your evaluation of a particular system. Thus, the most logical solution to measurement would be to test all communicators under consideration under the same conditions, and using the same material. In this way, a relative ranking of the systems can be obtained. It cannot logically be said, however, that one communicator is "15% better" than another, because the results are simply ranked data, not absolute in terms of percent intelligibility, etc.

Space does not permit an exhaustive discussion of factors involved in obtaining intelligibility scores; however, we can list a few of them such as word familiarity, syllable length and complexity, word use frequency, redundancy in sentences, talker and listener differences, and closed vs. open response sets.

Attempts have been made by myself [23] and several other people [24,35,45,46] to sort out the specific effects of air and HeO_2 mixtures on intelligibility. Most of us have used closed response sets, or gone to an exhaustive matrix assuming any response was allowable. The results of all the studies put together, frankly, are rather non-productive. There appears to be a rather strong trend (for both air and HeO_2 mixtures) for the manner of articulation of a particular phoneme to be preserved, with indications that the place aspect of phonemes is distorted. The voicing feature appears to be relatively stable. In work with normal air mixtures, a tendency for medial phonemes to become peripheral was noted, although at depths below 200' the changing acoustic coupling and transmission losses through the human pharyngeal wall appear to wipe out this effect; as a matter of fact, the tendency is for peripheral phonemes to be heard as medial, such as /1/ for /b/, /t/ for /k/, etc.

Specific intelligibility scores for various communicators are available in the literature;[23,24] from the success or the lack of success we have had with attempting to establish some degree of order with respect to intelligibility testing, several conclusions appear to be safe to state.

First, it is not likely that any communication lexicon based on phonemic restriction will be successful—relationships between phoneme classes appear to change as a function of depth and mixture.

Second, it is desirable to test communication devices with groups of naive, but live, listeners, using a closed set of test sentences or words.

Third, great caution must be taken in using multiple forms of the same test, under the assumption that "equivalent word lists" are in fact equivalent at depth. (Closed sets of words and sentences would greatly reduce this problem.)

Fourth, in actual operations, a restricted lexicon is needed, to increase the probability of a particular word occurring.

Fifth, a lexicon which includes certain types of words in specific positions within a transmission is needed.

Finally, the most obvious single need in evaluating communicators is to standardize presentation, conditions, depths, etc., in order that intelligent decisions can be made concerning the relative merits of the systems under consideration. Expensive and time consuming live talker/listener tasks in a standard setting come as close to operational evaluation as any system I can think of, and are infinitely better than submerging a group of divers with different systems on, and having each one count to ten.

4. REQUIREMENTS OF THE IDEAL HELIUM SPEECH COMMUNICATIONS SYSTEM

by J. H. Elkins

INTRODUCTION

Diver communications have been needed since man first ventured beneath the sea. The early diver was dependent upon hand-line signals from the surface and for many years this method was the only means available. So work on a limited piecemeal basis was started on underwater communications. But, before we really ever understood the problem and certainly before adequate equipment was developed, we were face-to-face with another problem. The breathing of exotic gas by deep divers offers unique problems in voice communications. Consequently, we began solving the second problem without solving the first. At this point I would like to back up and review the progress in underwater communications over the years.

The communications problem for the hard-hat diver was solved to a very limited degree in 1925; however, the advent of self-contained underwater breathing apparatus (SCUBA) brought with it a completely new set of problems. The diver was no longer speaking in an air environment (helmet), there was no tether to the surface, and to further compound the problem, the diver had a mouthpiece which filled his mouth and virtually precluded voice communications.

Various attempts were made to provide voice communications by the use of throat microphones, bone conduction pickups, etc., but in the final analysis they all provided little improvement over no communications at all and it was finally recognized that the diver could not talk with the mouthbit in place. This factor led to the design of a full facemask.

The full facemask, and later the oral-nasal facemask, was an important step in the right direction. The first full facemasks improved communications but also generated physiological problems in the form of carbon dioxide poisoning due to the large "dead air space" within the mask. The oral-nasal mask was designed to reduce this dead air space, and accomplished this objective to some extent.

The communications equipment for the oral-nasal mask (AN/PQC-1) was designed along the lines of surface communications equipment and became the first piece of underwater communications equipment designed for use by the U. S. Navy. This equipment was a masterpiece of packaging for this era (1957) but provided an intelligibility of less than 14 percent. There were three factors contributing to this low intelligibility. The mask used was the first full facemask and had a poor acoustic response. The bandwidth was even less than used in the common telephone (which is insufficient) and the microphone was a carbon button microphone very similar to that used with World

War II radio transmitters, and noted for its distortion. These factors added up to a poor communication system.

The second underwater communications equipment for fleet use (1962-64) was the AN/PQC-1A. This equipment was designed in the early days of transistors and was essentially a transistorized version of the AN/PQC-1. The same microphone was used, the same bandwidth, a slightly improved mask, physiologically, but since transistor technology had not advanced sufficiently, this equipment proved to be little better than its predecessor.

Since the unsuccessful AN/PQC-1A, there have been a number of commercial items which have shown promise.

Aquasonics (now Hydro Products) was responsible for a series of equipments operating on a frequency of 42 kHZ, amplitude modulated, which provided intelligibilities of approximately 70 percent. This equipment was limited in range (300 yards) and it made use of the NAUTILUS mask which is physiologically unacceptable. The reliability of this equipment was poor due to excess cabling from the diver's belt to mask and hood, which was subject to failure.

Since these developments, many attempts have been made to solve the communications problems of the diver; but few have had the background, experience, and intimate knowledge of the environment and technical know-how necessary to succeed.

One of the common mistakes is for a designer to conceive a new method of transmitting the signal through water, design the equipment, and then begin to look for a microphone and earphone with which to use the new equipment. A search reveals no microphone or earphones specifically designed for use under water, so the designer resorts to one of two possible approaches. A conventional air microphone will be used but "modified" for use under water. The modification will consist of a waterproof covering, which destroys the sensitivity and frequency response so vital to successful underwater communications. The other approach, equally unsuccessful, entails the decision to develop a microphone for use with the new equipment. The bandwidth chosen will be compared with what is required on telephone circuits and will not be sufficient for a diver at depth. If the designer happens to be correct in bandwidth selection, then the pressure to which the microphone or earphone is subjected, will either cause the microphone/earphone to become insensitive, change the frequency response, or become inoperative.

Some equipment that has been developed on the commercial market has achieved some degree of success as mentioned earlier and hard-wire communications developed for Sealab II and III by NCSL have achieved intelligibilities of 80 percent (at shallow depths) and could be adapted for other tethered applications.

Now, back to the more recent problem of helium speech. To avoid nitrogen narcosis, divers breathe helium-oxygen at depths of 200 feet or more, with a resultant upshift in voice frequencies to the detriment of intelligibility. Down-shifting the frequencies has been attempted by means of various unscramblers with varied success.

High-voiced frequencies when breathing helium-oxygen result because formant frequencies of helium speech are related to those of normal speech in air by the ratio of sound velocities in the two mediums. The velocity of sound in dry air is about 331 M/SEC and in helium is about 1284 M/SEC.

Diving to depths of 600 feet is routine today and we have every reason to expect that depths of 1000 feet and beyond will become routine within a few years. Based on these projections then the helium speech problem is likely to become even more acute in the near future.

Some helium speech unscramblers tested by the University of Florida in May 1970[24] attained scores of only 45 percent. This emphasizes the need for more work in this area. There are various theories which attempt to account for the somewhat limited success in helium speech processors. In my opinion, the largest contributor is the lack of knowledge pertaining to the microphone requirements with which to feed the unscramblers. One early unscrambler had circuitry incorporated which actually limited the passband to an upper limit of 5000 HZ. The natural tendency is toward a flat microphone regardless of the bandwidth but inside a facemask a flat response is not what is required but rather a 12 DB/octave upslope on the high end.

REQUIREMENTS

Diver. One requirement concerning underwater communications, which is totally unrelated to equipment, but which is just as vital, is the qualifications of the diver who wears the equipment. Diver experience is a large contributor to successful communications. Some of the new masks are frightening to the inexperienced who find it difficult to make themselves understood wearing these masks regardless of the equipment under test.

Input. Electronically speaking, one of the first requirements for helium speech processors is a good quality "front end," that is, the electronics which precede the unscrambler and determines the quality of the input data. Many of the early tape recordings of helium speech which were used as a basis for equipment design, were of notoriously poor quality. They were made with poor microphones or microphones of unknown characteristics connected to unshielded cables and recorded on recorders of questionable quality. So we might say that the early developer was doomed to failure before he began. Today, however, the situation is very much improved as you will learn from Dr. Morrow's talk which follows my own.

Bandwidth is a vital parameter that is often overlooked. We are no longer speaking in the range 200-3500 HZ as with the telephone, but rather more like 300-10 KHZ. Actual sonograms of vocal data taken at a depth of 460 feet makes it obvious that an upper limit of at least 10 KHZ is required. It is of interest that the word "team" was reconstructed by use of an unscrambler indicating that there is still progress to be made.

A good quality mask and a good quality microphone will not necessarily make a good communication system unless they are compatible. For example, a Bruel and Kjaer standard microphone frequency response as it is in a simulated free field is very flat but when put inside a facemask, it deteriorates drastically. The mask and microphone must therefore be considered together and not individually. Some microphones deteriorate when subjected to the deeper depths. Still others show different characteristics in different masks and marked differences are noted between helium and air. These differences are, however, minimized in the case of the helmet.

Tests to date indicate the microphone developed by Dr. Morrow (LTV microphone) [32,33] compared to two other types is superior and that we can now start out toward a complete unscrambling unit with a quality microphone.

Conversion. Real time conversion is mandatory as a practical means of communications. That is, using data taken with the LTV microphone,[32] reasonably intelligible speech can be obtained by simply slowing down the tape recorder. This is an impractical method of speech conversion but it does prove, however, that all of the important frequencies are detected by the microphone.

Output. Hearing thresholds vary a great deal, especially among divers, which dictates that the audio levels should be somewhat higher than normal and adjustable.

Size. The unscrambler should be small and compact such that the diver can carry it and be provided with a corrected version of his own voice which has been proven to increase intelligibility due to diver adaptability.

Intelligibility. There are as many means of measuring intelligibility as there are people in this room. Intelligibility scores alone are meaningless unless accompanied by all the conditions under which the scores were obtained. I will not attempt to discuss all of the various methods of intelligibility measurement but will only describe the way we make these measurements at the Naval Coastal Systems Laboratory in Panama City, Florida. We have recently gone to the modified rhyme test (MRT). This method simplifies the taking of underwater data in that the diver does not write what he hears but chooses one of the six possible words on his list. These lists were statistically designed and computer randomized for our ongoing effort in underwater communications under the sponsorship of the Naval Ship Systems Command and others. We have been using this method, or phonetically balanced word lists, for the past eight years. At this point, the MRT method appears to be more practical.

Controls. The controls on helium speech processors should be minimized. Often the diver has his hands full just staying alive. An on-off switch and volume control should be all that is necessary. Any adjustments to compensate for different depths should be made either prior to the dive or automatically by pressure sensitive devices.

Power. The power required should obviously be minimal since the diver is carrying the speech processor. It should be possible through today's technology to combine the speech processor with other communications equipment and in fact, this is the approach that we at the Naval Coastal Systems Laboratory in Panama City are taking in our ongoing effort in communication.

When this system is completed about a year from now, we hope to provide the military diver with a modular communications system whether he is tethered or untethered saturated or unsaturated, at any depth.

5. THE INITIAL SPEECH TRANSDUCER AND ITS ENVIRONMENT*

by C. T. Morrow

The majority of microphones that have been used in deep dives are standard communication microphones with a response to only about 3 or 4 KiloHertz (kHz), which is not improved by a deep submergence atmosphere. When typical "high fidelity" microphones are used, their high frequency response is degraded by the deep submergence atmosphere. Taped helium speech recordings at about 400 feet supplied by Captain George Bond and by Harry Hollien of the University of Florida proved to be unintelligible to the novice listener. Sonograms showed no energy above 3 kHz. If one considers the upward shift of the resonance regions that occurs in helium speech, the feasibility of using a standard microphone becomes even more discouraging.

The purpose of our program was to develop improved speech communication in diving masks used in shallow and deep submergence atmospheres. As part of the program, construction of an experimental microphone with a response to 10 kHz was undertaken. The experimental models were conceived with the aid of A. J. Brouns. The initial gradient microphone contained a curvilinear aluminum diaphragm coupled at its center to a bimorph bar as in Figure 1. Later revisions have utilized a plastic dome coupled to a Bimorph ring. By compromising a certain amount of sensitivity, the mechanical resonance can be set near 20 kHz, which is out of range for speech even in helium atmospheres. By keeping the back of the microphone open, it becomes a gradient microphone. It is unlike most microphones which have a diaphragm resonance in the speech frequency range and, consequently, have two sets of parameters which affect the response. The first is the mechanical properties which don't change with depth; the second is the properties of the cavity which do change with depth. Thus, even if a microphone is "high fidelity" when you start with it, it is not when you get down to depth.

Upon completion of the construction of the experimental microphone, the next step was to measure mask cavity acoustics and response of the microphone in simulated mask cavities. To do this, a model cavity having a constant diameter but variable volume was built using plastic rings and an endcap which could be coupled on to a basic mouthpiece, and also a large rectangular plexiglass cavity.

An acoustic impedance calibrator was constructed using a Bruel and Kjaer (B&K) artificial voice, a Sierra Engineering 50th-percentile male anthropomorphic head, a sound transmission tube with a small amount of lamb's wool to decrease standing waves, and a sintered bronze disk (to provide a high acoustic source impedance at the lips) just under 3/4 inches in diameter. A 1/4-inch B&K condenser microphone

**References 32 and 33 treat this subject more comprehensively.*

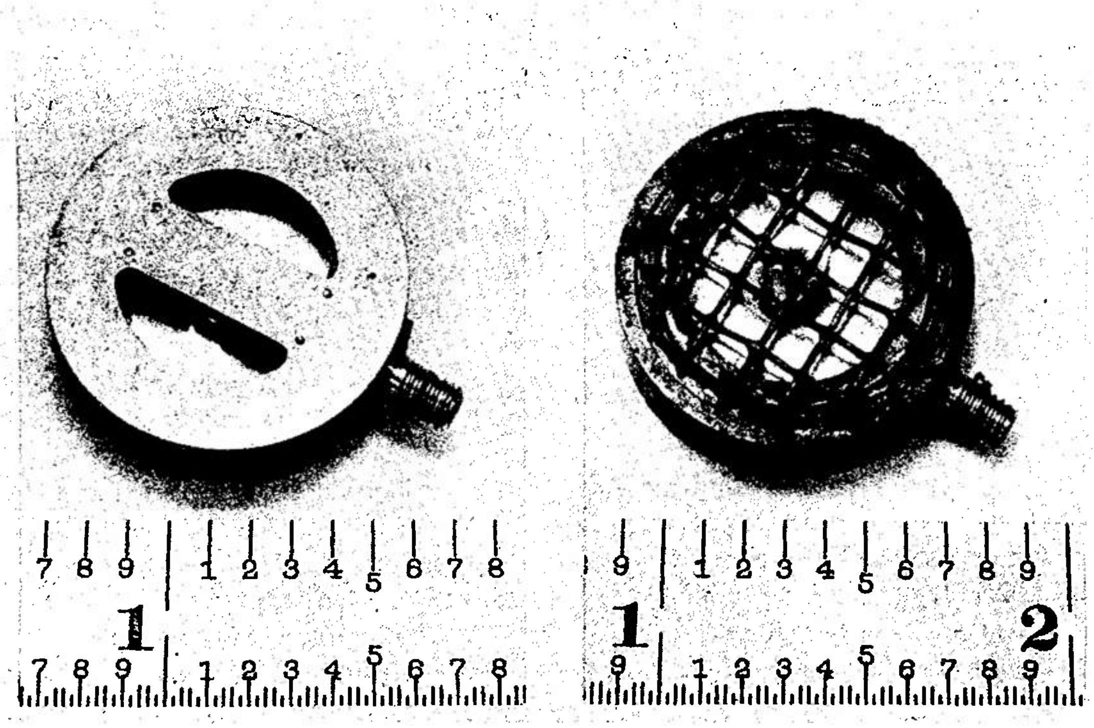


Fig. 1. (Morrow) The initial experimental microphone.

was mounted in the tube, just back of the disk, and was connected in an automatic gain control circuit to maintain constant sound pressure on the inside surface of the disk. A second 1/4 inch microphone was used to measure the sound pressure immediately in front of the disk. The masks and cavities tested were remotely positional on and off the head by a motor and lead screw in the base - a convenience when tests were to be performed in a helium pressure chamber. Pressure response curves for typical experimental cavities, as measured on the calibrator to indicate their acoustic impedance relative to open space, were obtained. According to our initial thoughts, the 17.4 cubic inch cavity would be close to an optimum for a mask without absorption, and preliminary listening tests with human speech appeared to confirm this. Except in the case of the 63 cubic inch rectangular plexiglas cavity, which had resonances in its front and rear walls, pressurization with helium shifted the curves upward in frequency essentially in accordance with the change in speed of sound. Sound absorption was effective in eliminating standing waves, but this was found to be unnecessary for intelligibility when the gradient microphone was used close to the lips.

Measurement of transmission to the outside as well as measurement of acoustic impedance was undertaken in order to understand the combined effects of cavity and transducer more fully. In the case of diving masks and other closed cavities, there is no useful radiation to the outside, and communication becomes dependent on a microphone that is not limited to instrumentation types. The relative transmission of a given microphone in a given location in a given cavity was obtained from its response to the calibrator, with and without the cavity. The response of the gradient microphone at the lips appeared simpler and more attractive than those of the pressure microphone, with no need for equalization for bass boost from the mask cavity and with promise for somewhat better intelligibility. When close to the lips, the pressure microphone showed a transmission curve similar to the impedance curve for the same cavity, as might be expected. The shape was, however, obviously different for other pressure microphone locations in the cavity.

Some recordings of a diver's speech were made at simulated depths of 650 and 460 feet at the Experimental Diving Unit in Washington, D.C. A ceramic gradient microphone with its preamplifier was used. The preamplifier had a 6 dB/octave high frequency preemphasis above about 1.5 kHz as a partial compensation for decreased intensity of high frequency harmonics in the voice in deep submergence. The microphone diaphragm resonance was at about 20 kHz, requiring no acoustic compensation which would degrade the deep submergence performance.

Using a "high fidelity" headphone, the intelligibility of helium speech directly from the microphone, or from the playback head of the tape recorder, was disappointing. It was better than the recordings supplied by Bond and Hollien, but not by very much. On playing the helium speech back at one-half of the recording speed, the articulation improved to an estimated 90 percent or better. In consideration of the generally poor reports on this method of translation, and the rather poor intelligibility of the untranslated signal from the new microphone, the effect was startling. In contrast to the half-speed playback, playing the tape at normal speed through an "unscrambler" of rather early design did not provide intelligible speech for any adjustment of the unscrambler. Within the reproducibility of the human voice, it made no difference whether the microphones were held in front of the lips by a boom on a headset or by a mount in the Kirby Morgan Clamshell Helmet. (Of course, if the helmet had been submerged, there would have been bubble noise. However, the gradient microphone would have provided some discrimination against this.) Likewise, a change in depth from 650 feet and 97.5% helium to 460 feet and 98% helium had no effect.

These observations were confirmed by sonograms. The helium speech was played back at half-speed. The shapes of the formant frequency contours were very similar to those for sea-level air, for the same words spoken, but they were about 40% higher in frequency for the half-speed playback. They were also more sharply tuned. There is no obvious nonlinearity in the formant shifts, suggesting that Fant's estimate of 150 to 200 Hz for closed-lip resonance may be a little high.

Based on our information gathered during this program, there are several principles which could be or already have been applied to helium speech unscramblers.

Simply playing back the tape recordings at half-speed serves very well for helium speech to 650 feet but this cannot be done in real time. A tape head built into a spinning wheel, almost surrounding by a moving tape has been used successfully to stretch or compress the time scale of speech without changing its quality. The device simply omits brief intervals on the tape or reproduces them more than once. The same principle has been applied to shifting frequency without changing the time scale. The most successful embodiments have incorporated small electronic computers to shift the frequencies.

A second method is to divide the spectrum into frequency bands and heterodyne each downward. Heterodyning a broad band moves first formants downward more rapidly than second formants and might tend to compensate for pressure effects in nitrogen-oxygen or neon-oxygen atmospheres. For accurate compensation of an approximately linear shift by helium, a large number of bands would appear to be necessary. Translators available now are generally limited to two or three bands.

A third method is to use an analyzer and synthesizer with a frequency translation incorporated between the two. If desired, the fundamental frequency can be left unchanged while the envelope of the harmonics is shifted.

It is generally reported that unscramblers must be carefully adjusted for each voice. The results of our investigations indicate that there should be no need for such adjustment if the design is adequate to preserve or generate a satisfactory relative frequency relationship between formants. Further, if the percentage of helium is in the nineties, adjustments should not be necessary for depths. With helium speech, the amount of frequency division should not be critical. The tape experiments which we have done, plus other experiments in the shifting of playback speed make it evident that halving the frequencies would be satisfactory for normal ears. For divers with severe high-frequency hearing losses, it may be preferable to divide by three or more.

In conclusion, a type of translator is necessary for understanding speech produced in deep submergence helium atmospheres. In combination with this translator, an adequate microphone such as the one used in this investigation is necessary. The microphone discussed in this presentation performs as a gradient microphone to 10 kHz or better in helium. It is insensitive to mask cavity acoustics and it withstands pressurization, decompression and the marine environment.

sponding resonance is shifted to 610 Hz, as indicated by the vertical dotted line. In a similar way it is easy to find how

was easier to understand as it increases the velocity of sound and accordingly shifts up the resonance frequencies of

6. PRESENT AND FUTURE WORK IN UNDERWATER COMMUNICATIONS AT THE SPEECH TRANSMISSION LABORATORY IN STOCKHOLM

by J. Lindqvist

Our interest in divers' speech at the Speech Transmission Laboratory started in 1962 when the physiologist Bertil Sonesson from Lund demonstrated the nasal quality of speech recorded at 11 ata (300 feet). At that time a physiological explanation to the

resonance frequency is therefore limited by the shunting effect of the walls. When a vowel sound is produced the participation of the vocal-tract walls in the resonance system will shift, especially the first formant, upwards in frequency. This upward shift is more

frequency has to be corrected which in its turn implies a non-linear frequency transformation. When a helium-oxygen mixture is used, this effect of the non-rigid vocal-tract walls is much less pronounced and is in fact only noticeable

pling between the body wall and the lead and the existence of other wall shunts, like the soft palate and the sinus piriformis down to the trachea.

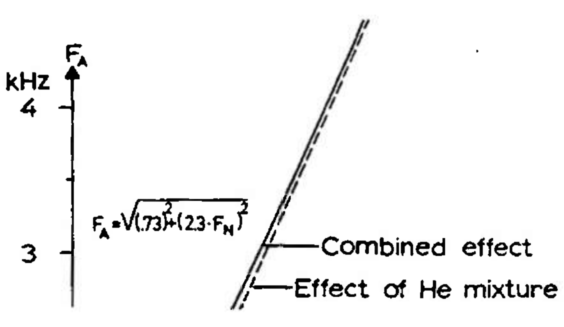
For a more detailed report on the ef-

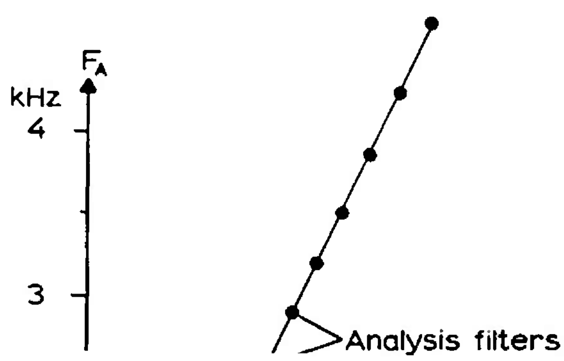
7. HELIUM SPEECH UNSCRAMBLER SIMULATIONS AT THE SPEECH TRANSMISSION LABORATORY

by Thomas Murray

Introduction. In 1969, we started to simulate Helium-Speech Unscramblers on our CD 1700 computer. The computer can store the speech signal to be processed on a CD 853 disc storage at a maximum sampling rate of 18 kHz, which gives a bandwidth of 9 kHz. Of course, we can store the data at double real time to decrease the needed bandwidth.

Figures 3 through 6 show actual formant frequencies F_A in different parts of a system as a function of normal formant frequencies F_N in air at sea level. To begin with all of them show the theoretical distortion curve for speech in 94% He and 6% O_2 at a depth of 660'. At this depth the pressure is 21 ata, and the lowest formant frequency for a certain diver is $F_W = .73$ kHz (Figure 3). The





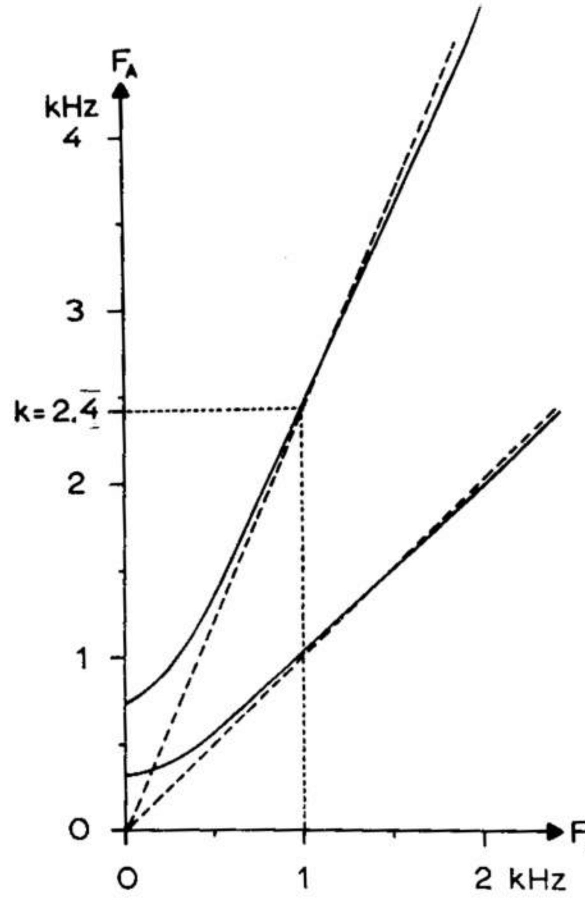


Fig. 5. *(Murray) Restoration of divers' speech with methods in the frequency domain.*

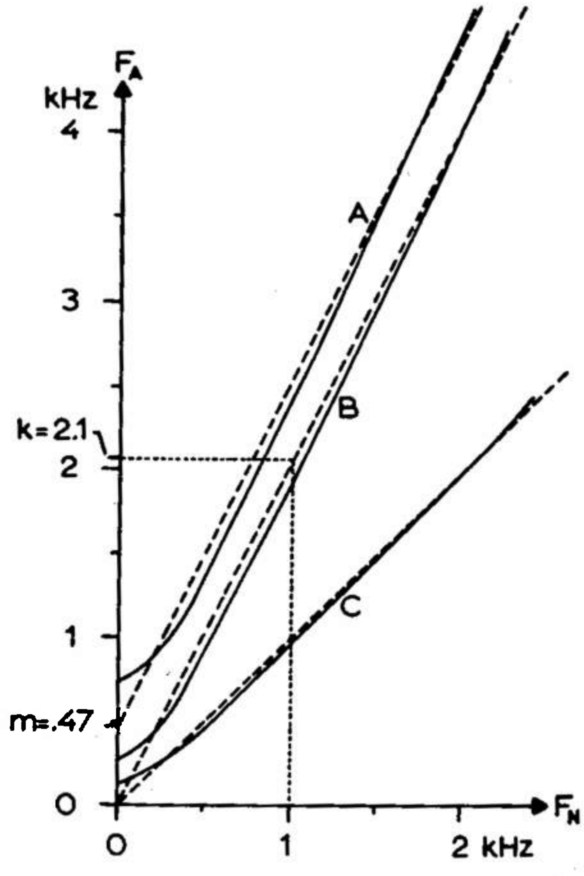


Fig. 6. *(Murray) Restoration of divers' speech with a combined heterodyne method and a time domain method.*

helium gas mixture causes a linear shift of the formant frequencies by a factor 2.3. The expression for the combined effect is shown in Figure 3. Both vocoder and time domain methods were simulated.

Vocoder Methods (VM). Two vocoder methods were simulated. The center frequencies of the synthesis filters were equally spaced along a technical mel scale. Both 11 and 16 synthesis filters were used (Figure 4).

In the Voice Excited Vocoder the rectified and smoothed signal from each analysis filter was modulated by a filtered and clipped speech signal. We tried to avoid the original formant frequencies from leaking through the modulator by: (1) the use of a very heavily clipped signal for modulation; and (2), modulating the different channels with signals, which are differently filtered before the final clipping.

In spite of these attempts, the original formants tend to leak through this type of unscrambler.

In the Pitch Voice Excited Vocoder, all channels are modulated by the same signal, which comes from a one-shot. In order to optimize the pitch-synchronism, a few methods of triggering the one-shot have been tested.

Time Domain Methods. In the Time Domain Method, TDM, micro segments of the divers' speech are stretched by the transposition factor k. To do this you have to throw away some parts of the speech signal. In spite of this, we have a feeling that there is quite enough redundancy of the speech signal to carry the essential information.

Figure 5 shows the effect of this method in the frequency domain. As you can see, the nonlinear pressure effect is only partly restored. A better reduction of this error is achieved if, by using a heterodyne method (HM), the formant frequencies are moved down in parallel before the TDM is applied. In Figure 6 you can see this combined method shown in 3 curves: Curve A shows the formants of the original divers' speech. Curve B shows the formants after the heterodyne. Curve C shows the formants after both the heterodyne and the Time Domain device are employed. As you can see, a better reduction of the pressure effect is achieved with this method.

Conclusion. We have not had the opportunity to compare these four methods on a very large amount of materials. Nevertheless, from the test results we have obtained to date, the Time Domain Method without heterodyne seems to give by far the best intelligibility, even when the other methods sound more natural. This and the fact that the TDM also is very useful for some other purposes like time compression and expansion, have encouraged us to design a hardware device based upon this principle.

8. DEVELOPMENTAL EVALUATIONS OF AND IMPROVEMENTS TO A HELIUM-SPEECH UNSCRAMBLER*

by R. A. Flower

Introduction. The basic objective of the work reported here was to achieve further improvements in the performance capabilities of the Singer-Kearfott Helium Speech Unscrambler technique. The unscrambler is an electronic signal processor which corrects on a real time basis the voice distortions arising from the use of helium in breathing gas mixtures for deep diving operations. A working model of the unscrambler exists in prototype form as a result of previously completed research and development activities.[12, 13, 14]

The areas under investigation include evaluation of a group of potential improvements in instrumentation, and evaluations by others of the prototype equipment. A description of these investigations and their current status is presented in the paragraphs that follow.

The evaluations of intelligibility reported herein were preliminary in nature, and intended only as a guide for deciding whether any of a group of possible modifications warranted incorporation in the Singer-Kearfott Unscrambler. These evaluations were based on judgment of experienced investigators rather than an objective listening panel methods. Results are therefore qualified by terms such as "subjective" and "apparent", and remain to be verified by formal test means.

AREAS OF INVESTIGATION

Compensation of Non-Linear Formant Shift. Previous studies[13] have shown that in helium distorted speech the lower formant frequencies are shifted by a greater ratio at a given depth than are the higher formant frequencies. For example, consider the predicted effects of depth on this pair of hypothetical formants:

	Sea Level	1000 Ft.	Ratio
F_x	400 Hz	1360 Hz	3.4:1
F_y	1600 Hz	4300 Hz	2.7:1

The ratio 2.7 is the asymptotic value of the frequency shift, which represents a fair approximation for all formants having a sea-level frequency of 1 kHz or more. Many unscramblers, including our unit, apply a single frequency correction ratio, usually one close to the asymptotic value appropriate to the operating depth.

This, therefore, produces under-compensation for the formants lowest in frequency, suggesting the possibility that an adverse effect on intelligibility may result. To investigate this we have devised and performed an experiment using recorded helium speech stimuli.

This report is based on "Annual Technical Report on Helium-Speech Investigations", R.A. Flower from the Singer Company to the Office of Naval Research (Project N00014-00-C-0387), June 1971.

The raw speech spectrum was divided into two parts by means of separation filters. The higher frequency part was processed by one unscrambler, and the lower frequency part was processed by a separate identical unscrambler. A diagram of the apparatus is shown in Figure 7.

The filters were of a commercial type manufactured by the Kron-Hite Corporation, Model 3500. These have a band pass characteristic, the upper and lower cutoff frequencies of which are adjustable over the entire frequency range of interest. The attenuation beyond cutoff is at a rate of 24 dB per octave. The two unscramblers have an adjustable input/output frequency ratio in the approximate range 1.4:1 to 3:1.

The glottal sync circuit of the unscrambler[13] operating on the low frequency band was used to control the reset cycle of both unscramblers. This prevented the possibility of having the respective reset periods unequal.

Various combinations of filter crossover frequency and unscrambler frequency ratios were tried, and compared subjectively with the performance of a single unscrambler covering the entire band. Among the dual-unscrambler combinations tried, the best intelligibility appeared to result with a crossover frequency of 2kHz (lower band 1-2 kHz, upper band 2-10 kHz), and with the lower band input/output frequency ratio about 1.25 times that of the higher band.

This agrees with what one could predict on the basis of the non-linear formant shift characteristic. However, the optimum two-band instrumentation did not result in significant improvement in apparent intelligibility. We conclude that a two-step correction to the non-linear characteristic is not promising. The potential of a multi-step or continuous correction remains an open question.

Unscrambler Artifacts. Ideally an unscrambler should produce at its output a voice waveform essentially

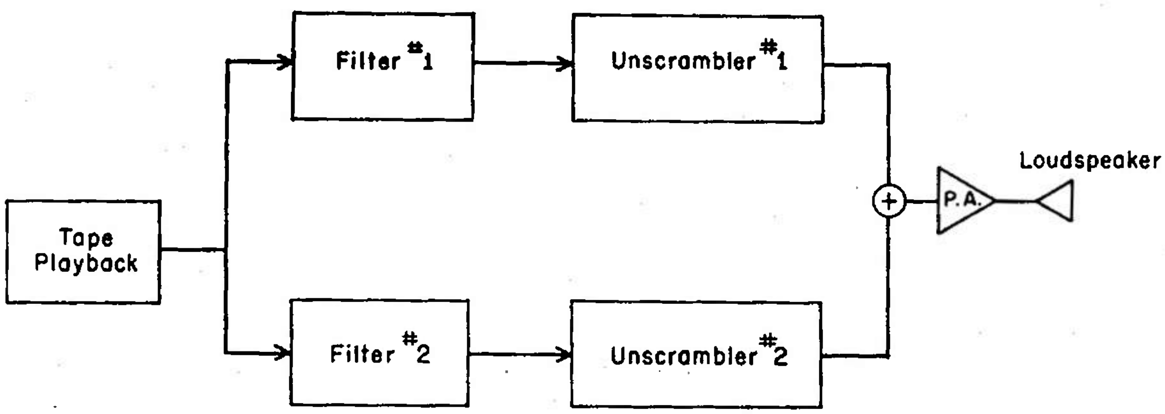


Fig. 7. (Flower) Diagram of the test apparatus for the non-linear formant shift.

identical to that of the same talker at sea level. Any practical device will, however, fail to reproduce perfectly all signal data appearing at its input, and in addition will add spurious components at its output. Such artifacts will, of course, vary widely in severity of effects on intelligibility. Some effects, such as noise and harmonic distortion, have been analyzed for voice communications systems in general, but others unique to the device in question are most readily assessed experimentally.

We therefore have examined on an empirical basis the major artifacts of the Singer-Kearfott unscrambler. These include synchronization sensitivity to signal polarity, significant levels of intermodulation components in the output, and voice wave envelope modifications.

Synchronization Polarity. The glottal frequency detector[13] or pitch synchronizing as originally designed is capable of generating a trigger on either an initial negative signal swing or an initial positive signal swing in each pitch period. This feature was incorporated so as to render the detector insensitive to the polarity (i.e., zero or 180° phase) of the voice signal, which is normally unspecified. Proper operation also is based on the assumption that the first half cycle of the lowest formant frequency is always larger than all succeeding half cycles in any glottal period.

Experience with a variety of sources of helium speech signals has shown that the pitch synchronizer performance is in some cases relatively sensitive to its amplitude threshold setting. This appears to correlate with exceptions to the assumption stated above, i.e., the first half cycle of the lowest formant is not always the largest in amplitude, whereupon the sync detector may alternate between the first and second, or even third, etc., half cycles, depending upon total signal amplitude.

When this condition obtains, the full synchronization capability can be restored if the sync detector is modified to operate on a unipolar basis. It then synchronizes only to an initial positive or an initial negative signal swing, depending on the selected polarity of the detector. Improvement in sync performance capability can therefore result with use of unipolar detection, but this requires the addition of a polarity selection control.

Intermodulation Components. Examination of the unscrambler output wave has shown evidence of spurious signals in the normal voice frequency range. These components add to the output noise background level, and thereby, tend to reduce intelligibility. The principal source of intermodulation products in the circuit under test was found to be high frequency signals appearing at the unscrambler input (voice outputs, circuit noise and acoustical noise), which heterodyne with the unscrambler sampling frequencies.

The problem is accentuated by use of high frequency equalization (higher gain at higher frequencies) at the unscrambler input. The solution is to apply a low-pass filter having a sharp cutoff at 10 kHz. The filter passes all necessary helium voice frequencies, and effectively excludes those components which lead to intermodulation products.

A four-pole Chebyshev filter was designed for this purpose. Subjective tests of the unscrambler with the filter added demonstrate significant improvement.

Voice Wave Envelope. The glottal wave in 1.0 atmosphere of air has an envelope that rises rapidly and decays to zero relatively slowly during each pitch period. The unscrambler output signal envelope differs in that the decaying envelope never approaches zero amplitude, because of pitch recycling[13]. A test was therefore developed to determine whether this has an effect on intelligibility.

The test circuit is shown in Figure 8. The unscrambler output is passed through a gain-control circuit having a very short time constant. The gain is varied at the pitch rate in accordance with the output wave of the function generator. With proper adjustment the unscrambled voice waveform can thereby be corrected approximately to the normal envelope characteristics.

Subjective listening tests indicated that the intelligibility was about the same with or without the waveform correction. We conclude that this artifact has no significant effect on intelligibility.

Automatic Gain Control (AGC). Optimum unscrambler performance requires that the input voice signal be maintained at a moderately fixed average voltage level. While variations over perhaps a 2:1 range are acceptable, we find that voice output levels of any individual, and the difference in average levels between individuals, can be

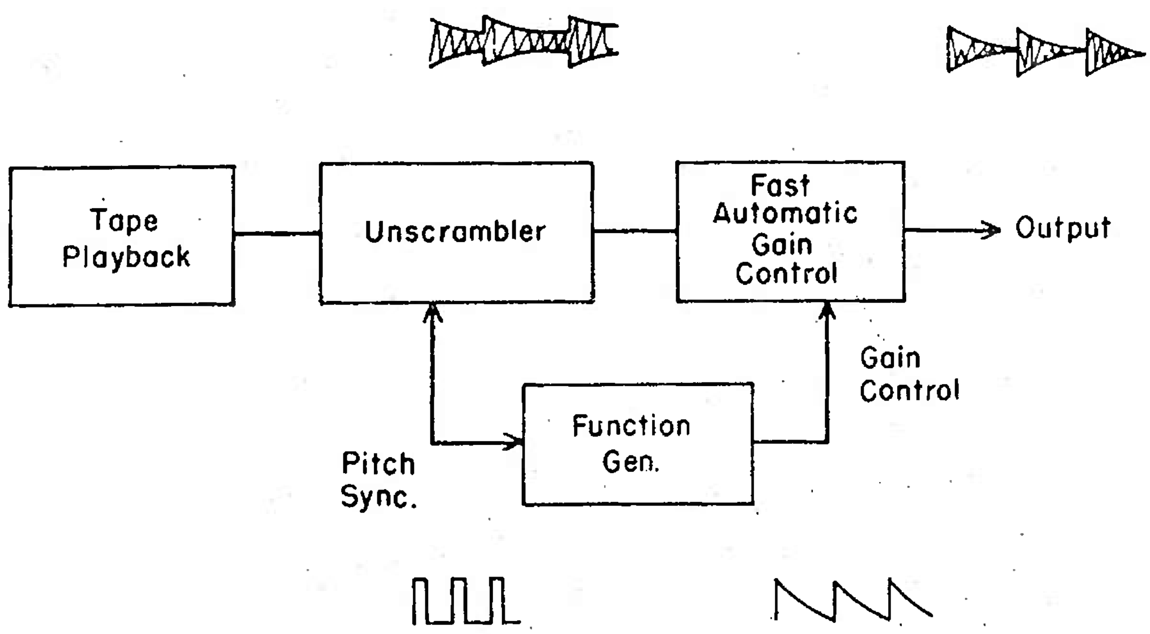


Fig. 8. (Flower) Schema of correction to acoustic envelope of helium-speech.

expected to exceed this range. As an interim expedient, we have been monitoring the voice levels with a VU meter and adjusting gain manually to maintain the proper input level.

For any ultimate solution, however, an automatic means is needed to control the input level. We have therefore examined the problem of providing automatic gain control, and have devised the following tentative specification:

Input Level Variation	30 db
Output Level Variation	3 db
Attack Time	0.01 sec.
Release Time	1 sec
Idle Gain	30 db below max. gain

It is expected that these values can be realized within the current circuit design state-of-art. However, we do not plan at the present time to pursue the design and fabrication of the circuit.

Low Quality Signals. We intended to investigate unscrambler performance with stimuli having relatively high background noise levels, but have found no recorded signals available with adequate parameter documentation. The plan is to record new material with the assistance of the Navy Experimental Diving Unit. A firm schedule remains to be worked out.

SUMMARY

The basic objective of this work was to continue evaluations of and achieve further improvements in the Singer-Kearfott unscrambler technique. Investigations show that improvements result from using single polarity rather than bipolar pitch sync detection, and from excluding input signal components above 10 kHz. Simple compensation of the non-linear formant shift, and modification of the glottal wave decay rate produced no apparent improvement.

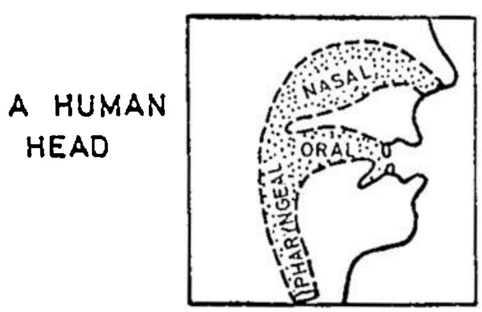
9. THE ADMIRALTY RESEARCH LABORATORY PROCESSOR FOR HELIUM SPEECH

by J. S. Gill

When diving to great depths it is necessary to breathe a mixture of oxygen and helium in which the partial pressure of oxygen is maintained at the same magnitude as in air at sea-level. Speech produced in this environment is badly distorted and at depths greater than about 600 feet, it becomes virtually unintelligible. The Admiralty Research Laboratory Helium Speech Processor was developed in 1969 to meet the Royal Navy's needs when diving throughout the range 0 to 2000 feet.

The principal cause of speech distortion in oxy-helium is the increased velocity of sound. Speech is produced by exciting the resonances of the vocal tract (see figures 9 and 10) by puffs of air from the larynx during voiced sounds and by turbulence at constrictions during unvoiced sounds. The intelligence is conveyed by varying the vocal resonances and excitation. Typically a male vocal tract is approximately 7 inches long and when relaxed, corresponding to the neutral vowel, and filled with air, the resonances occur at 500, 1500, 2500, etc. Hertz. The frequencies of these resonances depend upon the velocity of sound within the vocal tract. When, for example, a diver is speaking at 1500 feet depth in helium-oxygen these resonant frequencies are almost trebled and speech is completely unintelligible. The periodicity of the larynx excitation is virtually unaffected by the gas mixture, apart from the usual increase which occurs during stress.

Although some distortion is caused by high ambient pressure, for example the consonant-to-vowel ratios are reduced, by far the most significant distortion is caused by the effect of the gas mixture on the resonant frequencies.



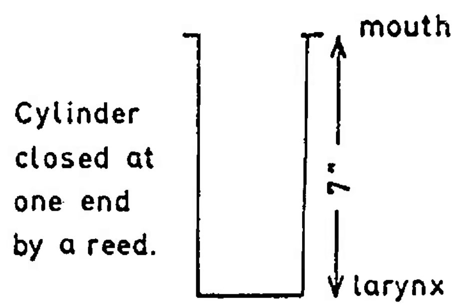
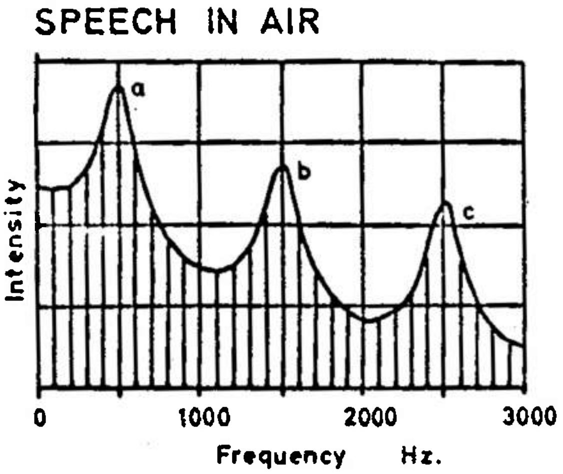


Fig. 9. (Gill) Diagram of a simplified model for a neutral vowel [ɛ].



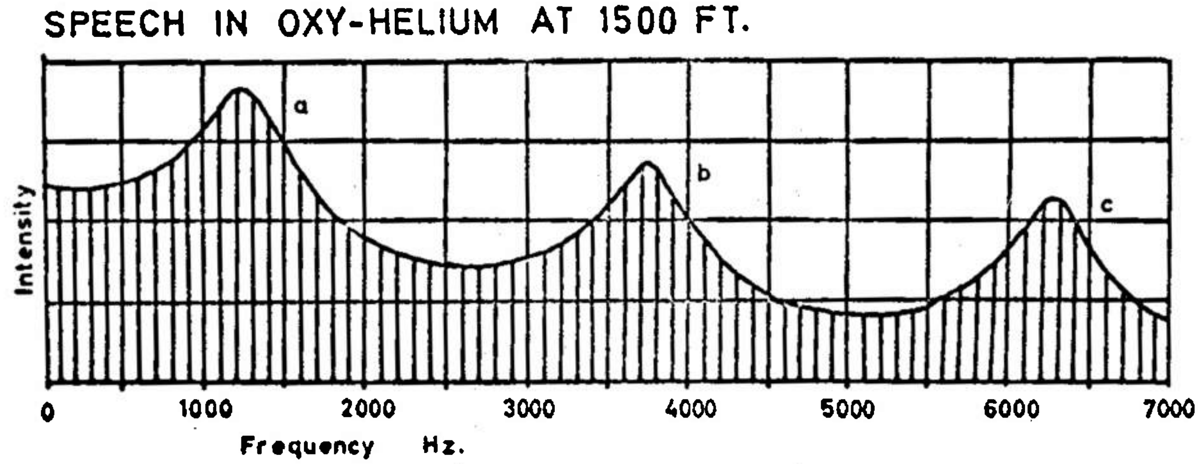


Fig. 10. (Gill) Acoustic spectra showing vocal resonances for a neutral vowel

Recordings

Before a processor could be developed, it was necessary to obtain good recordings of speech in oxy-helium at depth. We are indebted to LT J. Bladh, USN and Petty Officers Cook and Frazer of the Royal Navy who volunteered to dive to 800 feet for this purpose. This dive was carried out in the Deep Trials Unit at Alverstoke during August, 1968. A hydrophone and a moving-coil microphone were used and the signals were recorded on twin-track magnetic tape. Short-term spectrum analysis of these recordings showed clearly that the frequency response of the microphone varied with ambient pressure, whereas the hydrophone performed satisfactorily at all depths. The hydrophone was insensitive, having been designed to operate in water. This necessitated close-speaking, which improved the signal/ambient noise ratio. The hydrophone efficiency was greatest when operating at maximum depth, where the ρc of the medium was more nearly matched to the hydrophone. Care was taken to minimize the noise from the carbon-dioxide scrubbers.

Preliminary Studies

There was, by late 1968, a substantial volume of literature concerning helium speech processing but no known equipment met our requirements. The available literature concerning the application of band-shifting,[4, 27, 28] vocoder[22, 40] and time-domain[3, 39, 50, 51] techniques was studied and various possibilities were investigated by computer simulation.

The band-shifting technique was not adopted because it was not possible, with a small number of channels, to segregate and to control the frequencies and bandwidths of the vocal resonances and the resulting speech was anharmonic.

The vocoder method was simulated, in both voice-excited and pitch-extracted versions. Acceptable quality, albeit "vocoder quality", was obtained but this solution was not adopted because it would have involved substantial analogue circuitry and could not conveniently provide continuously variable frequency-compression.

Time domain processing appeared to offer the attractive possibility of a processor with a continuously variable expansion ratio which could be constructed using existing integrated circuits. There is a long history concerning possible applications of Doppler Scanning methods to bandwidth compression, for example French and Zinn[16], the German Tonschrieber (1939 to 1945), Gabor[18] and Fairbanks, Everett and Jaeger[8]. The first available reference to the application of these techniques to helium speech processing appears to have been Stover[50] followed by Brubaker and Wurst[3] and Westinghouse[51].

The first system to be simulated was based on the Gabor[18] technique using a pitch-synchronous scanning window approximately 50 milliseconds wide. This worked well but required a large store. The window width was therefore reduced to 4 milliseconds and the quality still appeared to be acceptable, even when the scanning was not pitch-synchronous. This tentative conclusion was based on the simulation of one short sentence only. Later, when the breadboard model had been built, pitch synchronism proved to be essential.

Principles of Operation

The basis of the method which has been developed at the Admiralty Research Laboratory[20, 21] is to write sections of the speech in a temporary store and then to read them at a lower rate (Fig. 11). During voiced sounds these sections are taken from the most intense part of each larynx period and the remainder is rejected. During unvoiced sounds the sections are taken less regularly and are more closely spaced. The frequency compression resulting from the lower replay rate is inversely proportional to the time expansion.

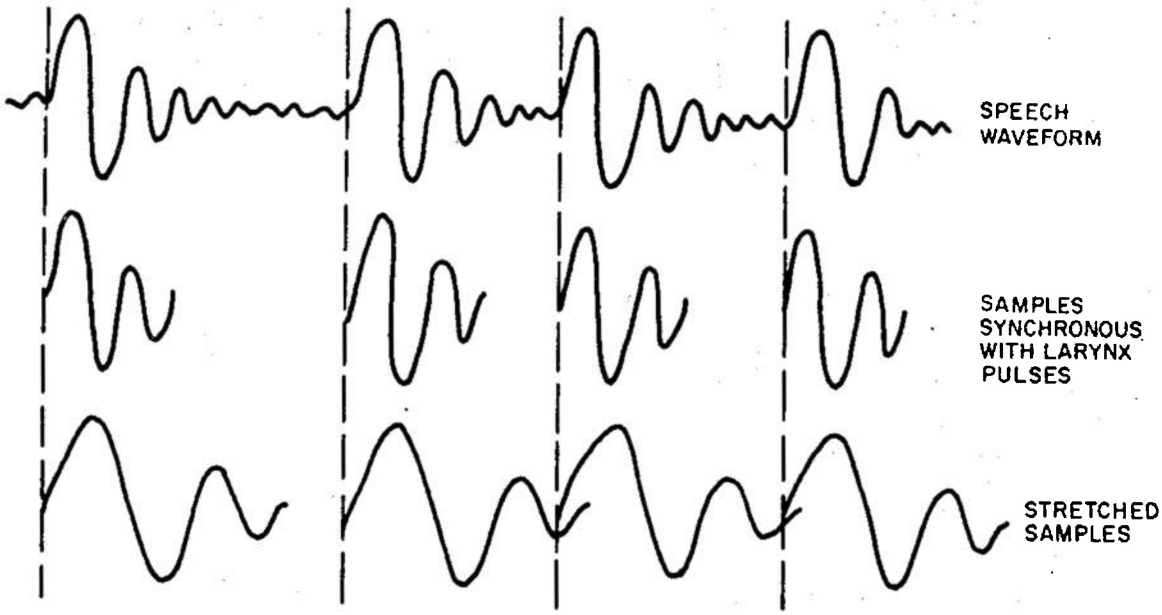


Fig. 11. (Gill) Waveform diagrams to describe the principle of operation of the Admiralty Research Laboratory Processor.

The section length must be less than the shortest larynx period, otherwise the larynx periodicity will be destroyed, but it must be sufficiently long to ensure that the first formant of the helium speech is adequately defined. In this application a length of 2.5 milliseconds was chosen to allow operation at larynx frequencies up to 400 Hz. The expanded sections overlap whenever the product of section length and expansion ratio exceeds the larynx period.

Four temporary stores are used and at any time one is being written whilst the remainder are being read. This enables the expansion ratio to be varied over the range 1:1 to 3:1 whilst retaining all of the sampled sections. Operation at larger expansion ratios would call for an increase in the number of temporary stores in order to maintain the whole of the 2.5 millisecond sections.

Longer sections, for example whole larynx periods, could be accommodated by increasing the lengths of the stores and incorporating fast-shifting between read-in and read-out, but this has not been necessary. The technique of using the most intense portion of each larynx period is advantageous when operating at a low signal/noise ratio.

The Processor

The system is outlined in Fig. 12. Signals from the transducer are filtered by a 12 kHz low-pass filter and then pass through an adjustable shaping network to the larynx pulse detector and analogue-to-digital converter. This

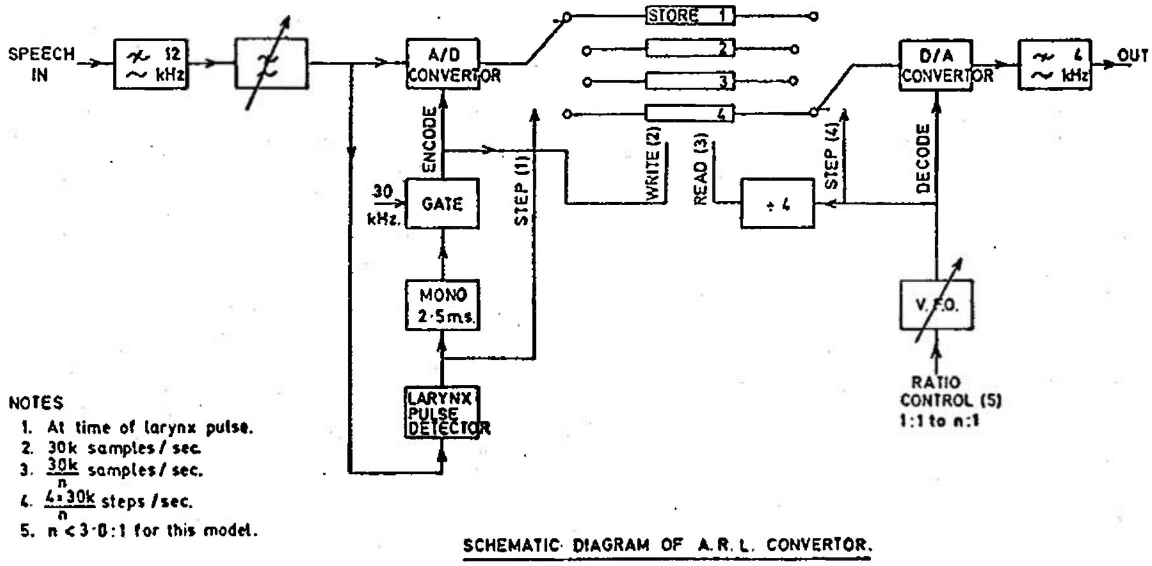


Fig. 12. (Gill) Schematic diagram of the Admiralty Research Laboratory Convertor.

adjustable network can provide up to 20 dB lift in the upper part of the spectrum to enhance the unvoiced sounds.

The larynx pulse detector selects the most intense 2.5 millisecond-long section of each larynx period. These sections are encoded at 30 kilosamples/sec into 8-bit PCM and read sequentially into shift-register stores. During voiced sounds each one of the four banks of storage contains the first 2.5 milliseconds of a larynx period. During unvoiced sounds the monostable fires almost continuously and most of the speech is stored.

Read shift pulses are applied to the 3 banks of storage which are not currently being written and these signals are interleaved and decoded at four times the read shift rate. The signals from the digital-to-analogue convertor are passed through a 4 kHz low-pass filter which averages the samples from the three stores and removes the unwanted products of the sampling process. The variable frequency oscillator which controls the overall expansion ratio is adjustable over the range 120 Hz to 40 kHz to provide any ratio which may be required within the range 1:1 to 3:1.

This model operated successfully throughout the record-breaking dive to 1500 feet, made by John Bevan and Peter Sharphouse of the RNSS at RNPL, Alverstoke, during March 1970. The processed speech was at all times highly intelligible.

Comment

The possibility of using analogue shift registers was contemplated at the breadboard stage of the project, but reliable digital components were readily available and were consequently used.

Improved methods of analogue storage have recently been developed commercially and at least one of these, the "bucket-brigade delay line",[41] may provide an alternative technique for a processor based on the above principles.

10. AN INVESTIGATION OF HeO_2 SPEECH UNSCRAMBLERS UNDER CONTROLLED CONDITIONS

by H.B. Rothman and H. Hollien

INTRODUCTION

One of the major obstacles to man's exploration of the oceans is the inadequacy of voice communications among divers and between divers and surface support personnel. It is well established that a talker in an environment of high helium concentration and under high ambient pressure experiences severe distortions in the intelligibility of his speech. Although the fundamental frequency of the talker's voice is not affected, there is an upward shift of formant frequencies (formant being defined as an area of acoustic energy maxima which specify vowels). This upward shift, which is nonlinear at low frequencies and is linear at high frequencies is one of the causes of distortions in the production and perception of speech.

Helium speech unscramblers, designed to improve speech intelligibility distorted by breathing helium/oxygen mixtures, have been developed. Generally, these systems attempt to process variously the grossly unintelligible speech resulting from the effects of HeO_2 breathing mixtures and high ambient pressure, and to reconstruct such signals in order to provide adequate oral communication. There are two general methods being used for HeO_2 unscrambling. They are: 1) frequency domain processing, where the spectrum of the signal is manipulated and 2) time domain processing, with the time-varying signal being manipulated. Some unscramblers use combinations of both techniques.

Several investigators[2, 19, 31, 43] have investigated the formant frequency shift of vowels in HeO_2 and found them to be close to predicted levels. Closer agreement between the shifts found and predicted would probably have been reached were it not for the fact that many of the tests were done breathing HeO_2 at atmospheric pressure. Therefore, changes due to pressure did not occur. While some of these studies[2, 50] provide much needed base-line information, they are of limited value since they have examined the formants of vowels produced in isolation. Other investigations by Copel[4] at the Naval Applied Science Laboratory (NASL) and by Brubaker and Wurst[3] of the Singer Company have only studied F_3 and F_1, respectively. The Singer group also investigated consonant-vowel amplitude ratios which are elevated, i.e., the consonant peaks are depressed in relation to the vowel amplitudes. However, no one has looked at transitions -- and studies of the acoustic characteristics of speech indicate that it is often the transitions between consonants and vowels that carry the relevant information for perception. For example, researchers at Haskins Laboratory[5] have shown that the direction and duration of transitions, particularly for the second formant, play a substantial role in the determination of perception.

It is obvious from the above that many different types of speech materials, recorded under a wide variety of conditions, have been used to obtain data on the distortions caused by HeO_2 and on the corrective ability of various unscramblers. Because of the disparity between the methodologies employed, it is difficult to compare the data in a meaningful manner. To correct this situation, the Communication Sciences Laboratory has undertaken to develop standardized procedures, as far as is possible, for testing the performance of HeO_2 speech unscramblers. To develop an experimental procedure of this nature, three sets of protocols are specified. First, the exact nature of the equipment to be evaluated is determined as far as possible; secondly, speech intelligibility tests are conducted, using standardized speech material equated for difficulty, and finally an error analysis is carried out.

This paper will present results on speech intelligibility based on the performance of HeO_2 unscramblers through 1970. The speech material used was collected primarily at the Navy's Experimental Diving Unit (EDU), Washington, D.C. on various occasions and under different conditions. This report will first present data collected "on-line" at EDU. Secondly, we will describe the development of our "standard" test for HeO_2 unscramblers. Thirdly, we will present data from our "off-line" evaluations (using the standard test).

On-Line Evaluation of HeO_2 Unscramblers

The first set of recordings were obtained during Sealab-3 training dives at EDU. The facility utilized may be seen in Figure 13; it consists of two large cylinders. The one on the right is a horizontal two-lock chamber and the other is a vertical two-level chamber with the bottom section extending well below the level of the horizontal cylinder. Thus, three of the four units are living, dry work (igloo) and "wet" work sections (the lower level of the vertical cylinder is usually filled with water); the outer chamber serves as an emergency "lock-in."

In order to make recordings of the divers' speech without excessive reverberation from the steel walls of the chamber, it was necessary to provide an area surrounded by acoustically absorptive materials. Since space considerations preclude the introduction of an "acoustically isolated chamber" into the already crowded habitat, environmental modifications were accomplished by using the fiber-glass-filled mattresses from a set of bunks to form an enclosure. This enclosure, with a fiber-glass-filled pillow in the rear and the talker acting as his own baffle at the front, served as a recording chamber.

To be properly conducted, assessment of system performance should follow set procedures; hence protocols were specified. They are as follows: 1) more than one (preferably 3-4) talkers should be used for each system and configuration because of the danger that talker variability could bias the results, 2) standardized word lists (equated for difficulty) should be utilized for much the same reasons, 3) all talkers should read at least two word lists via all unscrambler/microphone combinations, 4) each talker should read via a specific microphone through all unscramblers

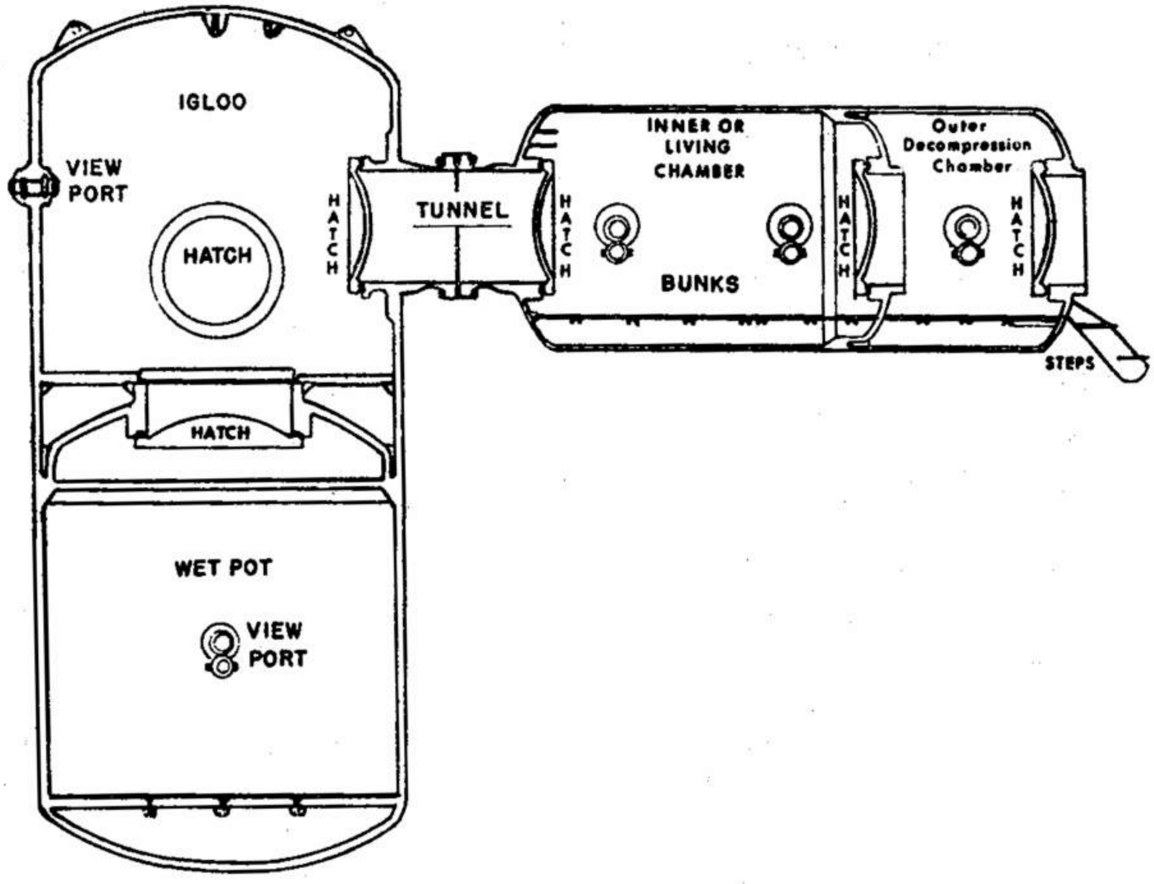


Fig. 13. (Rothman & Hollien) Schematic diagram of the hyperbaric facility at the Navy Experimental Diving Unit, Washington, D. C.

(and unprocessed) simultaneously (see Figure 14) at least 10 listeners (preferably 15) should be used to obtain the intelligibility levels.

Evaluation of Responses

Tapes of diver/talkers responses were spliced to allow three to five second intervals between words. These tapes were played to a minimum of ten semi-trained listeners; i.e., University of Florida students selected on the basis of 1) being native speakers of English, 2) having normal hearing and 3) being capable of performing the required listening tasks. Before hearing the tapes, listeners are required to score at least 92% on a screening test which included 50 words from CID Auditory Word List A-3 (Hirsh recording) recorded in +10 dB of thermal noise, 25 words recorded in a HeO_2 environment, 25 words from diver communication system recordings and 50 words from CID Auditory Word List 4-A. The final 50 words constitute the screening test.

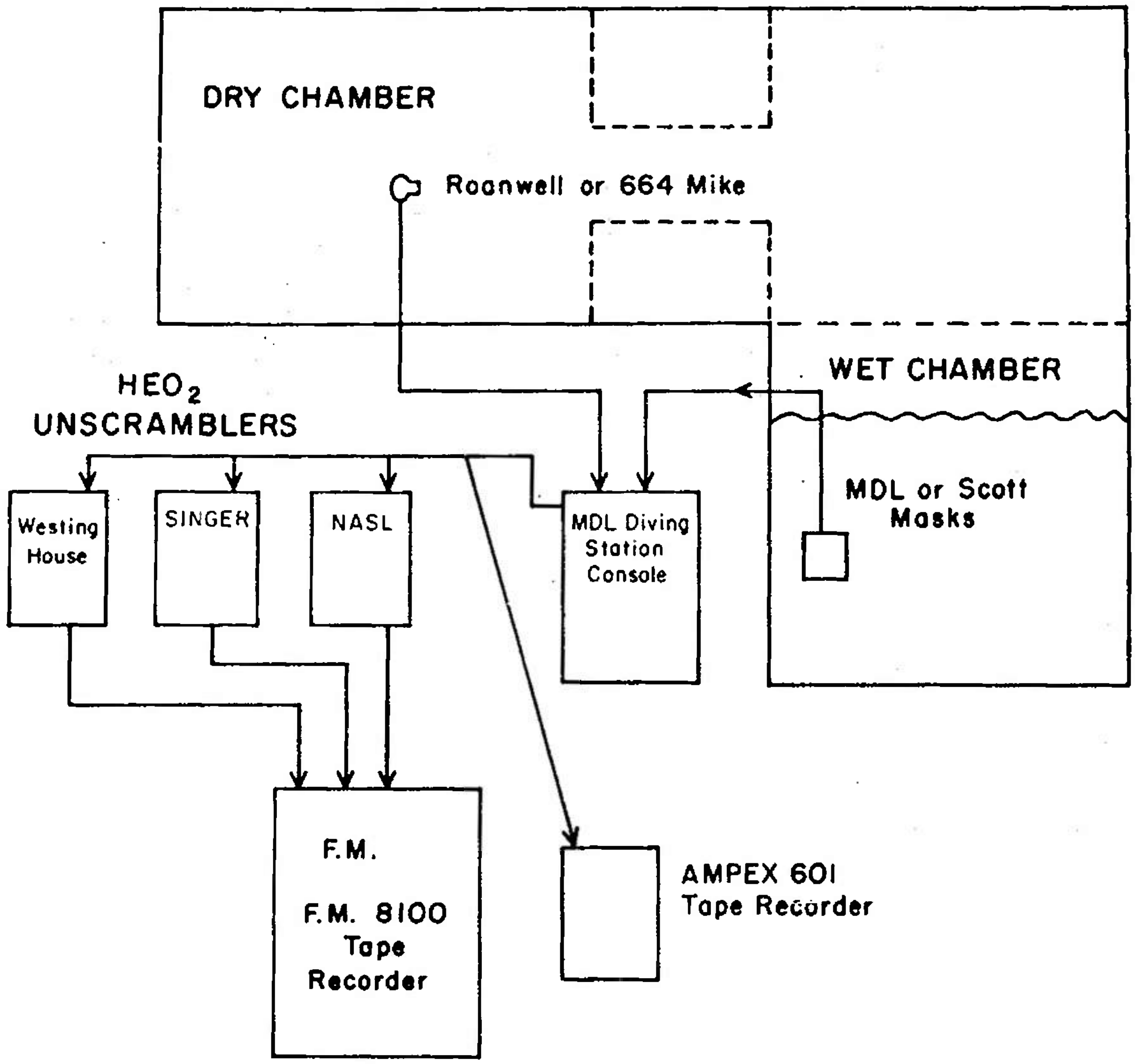


Fig. 14. (Rothman & Hollien) Schematic diagram showing the recording set-up used at the Navy Experimental Diving Unit. The outputs of the unscramblers were recorded on a Honeywell FM tape recorder and unprocessed speech was recorded simultaneously on an Ampex 601 tape recorder.

Each listener was asked to write down the words he heard from the tapes of divers' speech recorded from the output of the several unscramblers. Listener responses were scored for the number of words correct; the average percentage of words correct for each unscrambler was then used as its overall intelligibility score.

Table 1 provides the mean intelligibility scores for processed and unprocessed speech as well as a comparison of the Roanwell and Electrovoice 664 (EV-664) microphones. The Roanwell microphone was designed to be noise-cancelling, to operate under high ambient pressures and specifically to be the input transducer for the NASL unscrambler.

Table 1. Comparison of mean intelligibility scores for four diver/talkers in HeO_2 at 600 feet utilizing the Roanwell and Electrovoice 664 microphones. Recordings of unprocessed speech were made simultaneously with those processed through three unscramblers situated on-line. Each diver/score is the mean of four PB_{25} lists. N = at least 10 listeners in each case.

Microphone	Unprocessed	Unscramblers		
		H.R.B. Singer	NASL	Westinghouse
Roanwell	20.0	22.7	52.2	38.5
EV 664	7.5	10.7	27.6	27.0

As shown in Table 1, mean intelligibility scores for the EV-664 range from 5% for unprocessed speech to 27.6% for the NASL unscrambler. Mean intelligibility scores for the Roanwell range from 20% for unprocessed speech to 52.2% for the NASL.

From the data presented in Table 1, it is clear that under the operating conditions of this particular study, the Roanwell microphone proved to be a superior input transducer in all cases. It was expected to enhance the performance of the NASL unscrambler since it was designed specifically for that unit. However, it substantially increased the intelligibility when used with the other units. A mean intelligibility level of 52% cannot be regarded as satisfactory although it begins to approach a level whereby at least some intelligible voice communication can be expected between aquanauts situated in a chamber and the support groups on the outside.

In addition to the evaluation of the above unscramblers as recorded in a dry chamber, it was necessary to investigate their effectiveness using various input configurations such as would be used by a diver in the water. Obviously, the addition of a restricted cavity to the vocal tract produces profound acoustic changes in the resulting utterance; so does the interface of the diver's head with the water. This substudy was designed to add to the stockpile of information already gathered on the unscramblers and to examine the relative effect of two available diver's masks: the Scott and the MDL. Table 2 indicates that the unprocessed intelligibility for both masks is near the levels obtained in the chamber with the EV-664 microphone; the use of the Scott mask results in higher levels of intelligibility with and without the aid of the unscramblers, and, as stated above, the NASL unscrambler provides the greatest improvement.

Table 2. Mean scores for the Scott and MDL Masks at 600' in the wet pot at EDU. Each diver read two word lists on two different days. Means are corrected for unequal listener N's.

Mask	Unprocessed	HeO_2 Unscramblers		
		HRB Singer	Westinghouse	NASL
Scott	10.1	4.5	20.1	28.1
MDL	5.2	3.7	10.1	10.5

A second project was conducted to evaluate the performance of three HeO_2 speech unscramblers or processors used with five microphones. The unscramblers were those designed and fabricated by (1) Industrial Research Products, Inc., (IRPI), (2) the Raytheon Company and (3) Singer-General Precision, Inc., (G-P/Singer) (this unit is a newer and different model than the H.R.B. Singer); the microphones utilized in the evaluations included the following: (1) Industrial Research Products, Inc.; (2) Singer-General Precision; (3) Electrovoice 664; (4) U. S. Navy Mark-8; and (5) U. S. Navy Mark-II. Figure 14 is a schematic of the recording array.

Criteria 3 and 4 of the protocols mentioned previously were not met for this study, i.e., all talkers did not read at least two word lists via all unscrambler microphone combinations and each talker did not read via a specific microphone through all unscramblers (and unprocessed) simultaneously.

Table 3 presents individually, 97 obtained scores; each is based on at least 15 listeners. Table 4 is a summary table of the grouped data obtained from Table 3. It will be noted that the overall means are based on varying numbers of scores. That is, 30 lists were read for the IRPI unit and for unprocessed speech; 25 for G-P/Singer, 10 for Raytheon and only two for the NASL unit.

Rather extensive variability among the individual scores can be noted on Table 3. For example, scores range from 4.8% to 41.3% for unprocessed, from 5.9% to 65.1% via the IRPI unscrambler, and from 4.0% to 52.9% via the G-P/Singer. Variability is not uncommon to studies of this nature; it can be the result of such factors as 1) differences in recording procedure, 2) noise or improper grounding of equipment, 3) talkers and 4) word lists (such effects can be intermitent or cumulative). However, in this case, the variation among the individual scores in this evaluation are considered to be somewhat excessive. On the other hand, this talker variance can be explained to some degree; much of it seems due to the four

Table 3. Raw data from the evaluations of the four HeO_2 speech unscramblers. Scores are based on the percent correct responses from 15 listeners of each talker/microphone/list/unscrambler condition.

Microphone	Talker	List	Unprocessed	Unscramblers: IRPI	G-P/Singer	Raytheon	NASL
IRPI	Harder	T-3	21.0	48.8	-	48.6	-
		N-8	18.0	39.1	-	39.7	-
	Morey	O-8	26.6	40.3	-	-	30.7
		N-8	13.7	30.4	16.8	-	32.8
		M-10	-	-	-	54.4	-
		P-4	-	-	-	61.6	-
	Moore	P10	17.5	54.6	-	58.6	-
		T-8	18.0	47.5	-	57.8	-
MEAN			19.1	43.4	16.8	53.4	31.8
GP/Singer	Morey	O-8	19.7	29.6	18.3	-	-
		N-8	20.8	35.1	26.3	-	-
MEAN			20.2	32.4	22.3	-	-
EV-664-1	Harder	T-3	9.6	49.6	30.1	-	-
		N-8	18.1	48.8	15.0	-	-
	Morey	P-4	41.3	65.1	42.7	-	-
		M-10	36.0	56.8	33.9	-	-
	Moore	T-8	12.5	44.3	23.5	-	-
		P-10	24.0	35.5	18.7	-	-
MEAN			23.6	50.0	27.3	-	-
EV-664-2	Harder	T-3	7.3	13.1	16.0	35.1	-
		N-8	12.3	13.0	16.8	40.3	-
	Morey	P-4	7.1	17.1	22.7	-	-
		M-10	16.0	5.9	24.7	-	-
	Moore	T-8	9.7	-	9.6	22.7	-
		Q-4	4.8	10.5	15.2	-	-
		P-10	-	16.0	-	-	-
MEAN			9.5	12.6	17.5	32.7	-
Navy Mark 8	Morey	P-3	7.4	28.8	9.2	-	-
		Q-10	5.7	17.1	4.0	-	-
	Moore	T-4	18.6	41.3	15.7	-	-
		Q-10	13.1	17.1	11.7	-	-
MEAN			11.2	28.6	10.2	-	-

(continued next page)

Table 3. (Continued)

Microphone	Talker	List	Unprocessed	Unscramblers			
				IRPI	G-P/Singer	Raytheon	NASL
Navy Mark 11	Harder	O-3	5.1	-	22.3	32.6	-
		S-5	5.1	50.1	52.9	-	-
		Q-10	-	31.3	-	-	-
	Morey	S-5	16.8	33.4	22.7	-	-
		T-4	27.4	22.8	37.2	-	-
	Moore	O-3	6.2	15.7	18.1	-	-
		T-4	10.7	13.2	12.8	-	-
MEAN			11.9	27.8	27.7	32.6	-
Category Means			15.9	32.5	20.3	39.6	31.8
Individual Means			15.7	32.7	21.5	45.1	31.8

Table 4. Summary table of Unscrambler evaluation. Mean scores of words correct for each HeO_2 unscrambler and microphone. Diver/talker depth was 650' in HeO_2. N=15 listeners for each PB_{25} Campbell word list.

Microphones	Unprocessed	Unscramblers			
		IRPI	G-P Singer	Raytheon	NASL
IRPI	19.1 (6)	43.4 (6)	16.8 (1)	53.4 (6)	31.8 (2)
Singer	20.2 (2)	32.4 (2)	22.3 (2)	-	-
EV-664 (1)	23.6 (6)	50.0 (6)	27.3 (6)	-	-
EV-664 (2)	9.5 (6)	12.6 (6)	17.5 (6)	32.7 (3)	-
Mark-8	11.2 (4)	28.6 (4)	10.2 (4)	-	-
Mark-11	11.9 (6)	27.8 (6)	27.7 (6)	32.6 (1)	-
Category Means	15.9	32.5	20.3	39.6	31.8
Mean-all scores	15.7	32.7	21.5	45.1	31.8
Number talkers	30	30	25	10	2

conditions listed above, the balance to the sharp differences between the data collected by the authors (see EV-664-1) and those provided by others (see remaining data, but especially EV-664-2).

The main effects are apparent. First, none of the units improved speech intelligibility to levels that could be considered adequate for good communication. Hence, it cannot be said that any of them exhibit solutions to the HeO_2 communication problem. On the other hand, at least two of the units (Raytheon and IRPI) improved speech intelligibility by substantial amounts. Indeed, the Raytheon was the superior performer and provided nearly 200% improvement; the IRPI was second with about 100% improvement. With respect to the G-P/Singer unit, it can be said that it performed considerably better than did the previous Singer attempts but still not on the level of the other two units. Statistical evaluation of these data is difficult since the scores for the three units tested (omitting NASL) are not based on random trials; nor can the assumption of homogeneity of variance be met. Nevertheless, the differences are significant when a relatively insensitive test is used.

It is necessary, at this point, to comment on the lack of similarity between the levels demonstrated by EV-664 (1) and those by EV-664 (2). Since the talkers and lists were essentially the same, and the same microphone was utilized, it would be expected that the scores would be similar -- or due to adaption, that the second set would provide higher means (not lower) than the first. In carefully reviewing both sets of tape recordings, striking differences were apparent. Specifically, the recordings made by the authors (EV-664 (1)) are of substantially higher quality than those made later. The tapes in the second trial exhibited both greater variation in amplitude and more system and habitat noise. Hence, it is strongly recommended that whenever such evaluations are made in the future, the group that has the responsibility for the project should be allowed full control over all aspects of the investigation.

Comment also is necessary about microphone performance. The IRPI and G-P/Singer microphones performed at about the same level (even though the Singer scores are based on only two talker/lists, a tentative comparison is considered possible). Secondly, the EV-664, the Mark-8 and the Mark-11 also performed comparably. Finally, it can be noted that the two microphones in the first category (IRPI and G-P/Singer) provided mean scores that were about double those of the other three. It is suggested, therefore, that the G-P/Singer and IRPI microphones possibly will permit superior operation of HeO_2 speech unscramblers and should be considered for use in conjunction with such systems in future applications.

Additional information was obtained by comparing the three systems tested. Here, paired scores (same talker, word list and microphone simultaneously via the paired unscramblers) were used in order to obtain further direct comparison of the units. There were 23 lists allowing for comparison of the IRPI and G-P/Singer; four between the Raytheon and G-P/Singer and six for the Raytheon and IRPI. In these comparisons the Raytheon showed superior performance (100% better than the G-P/Singer and 29% better than the IRPI). In turn, the IRPI unit performed 39% better than the G-P/Singer device. In any case, the rankings among the three units were of the same order and of comparable magnitudes as they were on the other tables (main effects).

Development of an Off-Line Test

In the course of gathering data for the evaluation of helium speech unscramblers, the Communication Sciences Laboratory has collected a large inventory of speech materials representing a wide range of pressure and gas mixtures. Because of this extensive stockpile of speech material recorded in hyperbaric helium environments, it became possible to devise an off-line test for evaluating HeO_2 unscramblers. The advantages of an off-line test are obvious; further, such a test also is particularly useful for the preliminary testing of an unscrambler without the necessity of an actual chamber dive. The basic criteria used for selecting material for the off-line test were that it be 1) rigorous and 2) representative of the varied conditions found in on-line situations.

The recordings comprising the test are "good" recordings, i.e., they were closely monitored to prevent any unrelated distortions from occurring. However, the actual problems faced by the unscramblers have been carefully included. For example, noise is always a factor, so noisy tapes are included. Further, the diver will be at various depths in a hyperbaric chamber or the open sea, he will be breathing various mixtures of HeO_2 through different mask and helmet configurations while using different microphones under varying conditions of ambient noise. Hence, such conditions must be used and in this regard, the off-line test is intentionally rigorous and favors no particular unscrambler. It must be remembered that the results of an unscrambler's performance become especially meaningful when its performance is compared to the other unscramblers.

To be specific, the various word lists chosen for the test represent as many varied conditions as possible and allow for a rigorous evaluation of the performance of HeO_2 unscramblers. The criteria for the word list selection follows: (1) <u>High, medium and low intelligibility</u> - Intelligibility scores for unprocessed word lists for depths of 200, 450, and 600 feet that had been obtained from previous studies were divided into these three categories. One list each was chosen to represent high and medium intelligibility and two lists for low intelligibility were selected for each depth. All of these lists were first recordings made upon reaching depth: (2) <u>Noise</u> - Two recordings each, of word lists judged to be noisy, were selected for each of the three depths. The rationale for including this material is that a noisy habitat is a "typical" situation; (3) <u>Last recordings before starting ascent</u> (<u>LBA</u>) - These lists reflect diver intelligibility after he has a chance to modify his speech and has attempted to become more intelligible; (4) <u>Roanwell microphone at 600-825 feet</u>; (5) <u>Wet dive - 600 feet</u> - As all the above conditions occurred in a dry chamber, it was judged that this condition would add considerably to the evaluation of the performance of HeO_2 unscramblers since unscramblers will have to process divers' speech while they are working in the sea. In this case, diver/talkers wore either the Scott or MDL mask with the MDL microphone -- and had been at depth for some time when these recordings were made. The EV-664 microphone was used for conditions (1), (2), and (3). In all, 57 separate word lists are used in the CSL off-line unscrambler test.

Procedure

In order to conduct bench tests of unscramblers, recorded phrases were played on an Ampex tape recorder whose line output fed into the input of an unscrambler. The unscrambler output was fed into a Marantz amplifier coupled to a Marantz speaker. By such monitoring, we were able to keep the amplifier and speaker frequency response constant across all unscramblers. An attempt was made to "tune" the unscramblers according to the manufacture's specifications. After reaching this point, three listeners performed a modified method of adjustment to determine the adjustment which produced the greatest intelligibility. This was done by repeatedly playing a given signal while bracketing the area which gave best intelligibility. When agreement among the three listeners was reached, the unscramblers output was recorded on a second Ampex. This bracketing technique was carried out for each unscrambler, talker and condition. Input and output levels were carefully monitored to prevent distortion of the signal.

Results

Four unscramblers were tested for this particular run. They are Integrated Electronics Corp. (old NASL), the IRPI, Raytheon and Singer/G-P. As may be seen in the Table 5, the previous study showed the Raytheon unscrambler to provide the most improvement; the present study provides data which reverses this finding. That is, from an unprocessed score of 14.9%, the improvement provided by the Integrated Electronics Corporation unit is 24.6%; for the IRPI it is 31.7%; 20.0% for the Raytheon and 30.8% for the Singer/G-P.

Table 6 presents a breakdown of the data by depth and condition. A look at the scores for the Raytheon unit show that its performance deteriorated at greater depths especially when used with the Roanwell microphone and during the wet dive. A reexamination of Table 3 will show that the data from the unscrambler evaluation done at the Westinghouse facility, the high scores for the Raytheon unscrambler were primarily the result of its use with the IRPI microphone.

At present, we are continuing our off-line evaluation of unscramblers. We should be receiving tapes of our test through prototype units developed by Necton Bylinnium in Colorado, and by Gunnar Fant at the Royal Institute of Technology, Stockholm, Sweden. Other units we are hoping to receive data on include the Helle Engineering

Table 5. Intelligibility level for the four unscramblers tested during the summer of 1970. Means are not corrected for unequal listeners N's but have been equated for unequal number of lists read.

	Unprocessed	Unscramblers			
		IEC	IRPI	Raytheon	Singer
Mean	14.9	24.6	31.7	20.0	30.8
Number of listeners		755	781	785	788

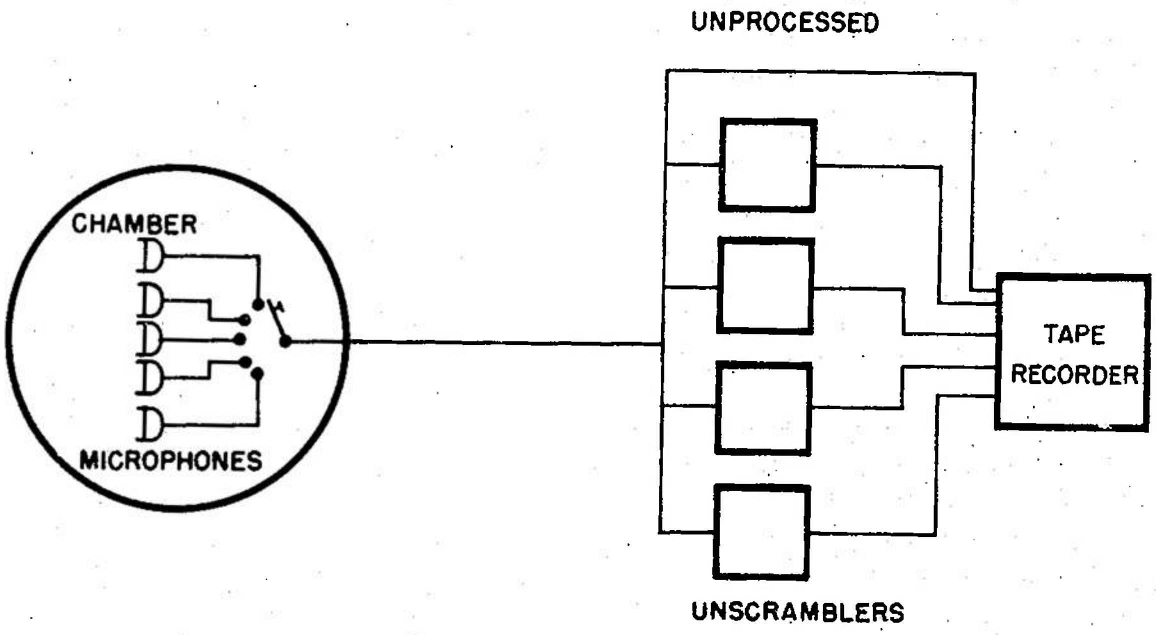


Fig. 15. (Rothman & Hollien) Schematic diagram of the recording set-up used at Westinghouse. This array enabled all microphones to be used with all unscramblers with simultaneous recordings made of unprocessed speech.

Table 6. Intelligibility levels (in percent) obtained for four HeO_2 unscramblers under the following varied conditions. Scores were not corrected for unequal listener H's, but have been equated for unequal number of lists used.

Depth	Condition	Talker/list Combinations	Unprocessed	Unscramblers			
				IEC	IRPI	Raytheon	Singer
200-450-600'	High, Med. & Low Intell.	12/12	22.8	31.5(156)	41.8(165)	31.0(174)	42.2(175)
200-450-600'	Noise	6/6	28.3	34.4(73)	43.8(87)	34.2(80)	41.3(80)
200-450-600'	LBA	6/6	36.6	31.8(79)	44.0(79)	28.7(82)	37.0(86)
600-825'	Roanwell Mike	8/18	3.7	*23.6(226)	**28.8(240)	15.1(252)	32.5(248)
600'	Wet Dive	7/15	7.9	13.6(221)	17.6(210)	8.0(197)	13.2(199)
	Overall Mean	39/57	14.9	24.6	31.7	20.0	30.8
	No. of listeners			755	781	785	788

* Only five lists were used.
** 13 lists were used.

Company's Hellephone, and the Standard Telecommunications Laboratory Ltd. unit from Britian.

Although all the data on all unscramblers is not in, the following generalizations can be made with confidence about unscramblers up to 1970. They are: 1) None of the unscramblers provide substantial enough speech improvement to allow for adequate diver-to-diver or diver-to-surface communication; 2) of the units tested the IRPI performed best overall; 3) the IRPI and Singer/G-P microphones appear to be similar in performance and superior to the EV-664, the Mark-8 and the Mark-11 microphones; and 4) whenever such evaluations are conducted in the future, rigorous and exacting procedures must be followed if valid performance levels are to be determined.

What then is the current State-of-the-Art? Obviously improvements will be made -- and are being made -- in HeO_2 unscramblers. Some of these improvements will occur because design engineers are becoming more aware of the complexities of speaking in the pressurized helium environment and are beginning to take basic research into consideration. We hope that our efforts in that regard and with respect to our evaluations of unscramblers will assist them.

Finally, a new aspect of our overall program is to study the human as a helium-speech decoder. Many divers report an increase in their ability to understand helium speech in a high ambient pressure environment after a period of time. Some of the data processors at CSL who have spent many hours listening to tapes of helium speech also report an increasing ability at decoding this type of speech even though it is greatly distorted. Because of these indications and because the human decoder should be a very efficient auditor, a promising area of research has opened. Hence, we now are developing techniques for identifying and training listeners to be efficient decoders of helium speech.

OPEN FORUM AND DISCUSSION PERIOD, CHAIRMAN's SUMMARY

by G. C. Tolhurst

The topics discussed during this portion of the workshop could be classified under six general headings. The subject matter tended to be additions and addenda to underwater communication problems generally, rather than those strictly associated with helium-speech translators. The six topics were: (1) a more detailed explanation of the project concerned with an universal underwater communication system development and evaluation; (2) the human talker-listener factors of a total communication system; (c) consideration in selection (if necessary) of diver-talkers; (4) possible methods of preprocessing the helium speech prior to translator processing which may enhance translator efficiency; (5) speaker-listener communication problems among pressure chamber subjects; and (6) factors inherent in quick solutions to operational problems.

NSRDL, Panama City Underwater Communication Project

At the present time navy divers cannot communicate reliably by voice at depths of 10 feet, or at any depth. A NavShips, SupSal, ONR project was initiated, using various components of underwater communication systems which had been developed and were available as "off-the-shelf" items in 1970, to combine these components into a useable system. Each sub-system component would be made electronically compatable so that any combination of three microphones, three facemasks, two helmets, several acoustic and wired transmission communication units, and five helium unscramblers could be tested and evaluated, both in bench tests, salt water tests, and psychophysical tests under laboratory and open-sea conditions. The goal is to achieve a design system readily adaptable to any and all types of diving situations from zero to 1000 feet depth and with a range (if acoustic transmission modes were employed) of 4000 yards. The resultant one or more "best" combination of components would become an interim model for Navy specified design and purchase. Any inadequate subunits of the total system should be revealed by the project testing procedures and yield indications for improved design and/or function. The detailed bench tests are almost complete, the sea test plans are written and should begin shortly. The completion date of the project is August 1972. Additional information can be found in NSRDL, Panama City reports, James Elkins, Project Manager or from Mr. Frank Romano, Naval Ship Systems Command, Arlington, Virginia.

Talker-Listener Factors of the Total Communication Chain

As pointed out by several of the workshop participants any speech communication network inseparably includes the speaker (encoder) and the listener (decoder) as well as any transmission link carrying the message. Questions were posed which inquired, generally, should there be attempts to train speakers and listeners and, if so, how and how much?

Since, in reality, there is no underwater speech transmission equipment that is operationally used routinely, the present training of divers consists of learning the code of tugs on the life line and the international system of hand signals used by SCUBA divers. In anticipation of functional voice communication systems, two research efforts were reported, one series from the University of Florida and another just starting with Westinghouse, aided by the University of Iowa. The data from the first series indicated that divers who were provided voice communication equipment took longer to perform a group of underwater tasks than divers who could not talk to each other. The unanticipated findings are confounded by many uncontrollable variables and are somewhat suspect, not the least of which was that the divers using communications systems were not trained in its use and were probably distracted by the noise and noise masked speech as well as the highly distorted speech. The Westinghouse study will attempt to train heliox speech listeners as a foreign language would be taught.

The question was posed by CDR Joseph Bloom; would talker-listener training be necessary since the preceeding two days of the conference seemed to show the transmission and translation of the speech signals were developed to the point that the "black box" between the talker and listener offered no further problems. In response, opinions were expressed which ranged from "training would be minimal" to "even under completely distortion-free conditions personnel need experience with a system in order to derive maximum utilization." Several concrete suggestions and rationalizations were given which would improve the talker or listener portion of the "communications chain."

However, several times during the general discussions and in subsequent "after session" talk it was iterated and reiterated that the apparent state of advancement in translator and microphone development was built upon evidence acquired under idealized or controlled laboratory conditions. The closest to realistic situations were in chambers simulating depth, rarely even in a wet pot, with little if any ambient noises of masks, valves, bubbles, etc. confounding the data. Research and development must take into consideration those variables. Additionally, only one method of helium-speech translation has been exploited and that is time sampling (of course heterodyne system was tried but most designers agree it is not a workable method, especially with the circuitry available one or two years ago). Other methods such as vocoder, digital coding-decoding or some hybrid of these methods with time sampling or heterodyning may yet prove to be more resistant to the operationally encountered speech distortion and masking sources. A viable program of research is still needed until efficient voice communications can enhance efficiently performed underwater work.

Diver-Talker Selection and Training

As mentioned above, the novice diver talker-listener will benefit from experiencing voice communications and the system operations in the environment in which he uses them. An analogy was given from data obtained from aircraft communications in which student pilots were given four hours of training in aircraft noise using the equipment they would encounter. On the average their intelligibility scores during training improved 15-20 per cent. Follow-up testing after some four years revealed they retained their training without loss.

In spite of the recent advancements made in diver communication system components, much research is still needed to determine what factors contribute to system integrity or failure when the components are combined into various configurations. There are but limited data concerning the effects of hyperbaric pressures upon the diver as a talker and none as a listener. Several of the participants raised the question of sidetone being introduced into a system, or not. The effects of spectrally distorted speech, time delay, intensity changes, etc. modifying the speakers' sidetone in air have been rather extensively explored and show the effects can either hinder or enhance talker intelligibility. The optimum sidetone manipulations for divers in a helium-oxygen breathing gas, under varying pressures, in various masking noise environments both in and out of the water need to be systemmatically studied.

It is possible, also, that no amount of training or equipment refinements will yield one hundred per cent transmission of voice messages. In such situations a diver lexicon may be needed in which communications will be limited in vocabulary and syntax. There are many examples of voice communication being possible only through the use of some standardized speech format, i.e., aircraft flight clearances, sound powered phone talkers, etc. Additional explanations were given by Dr. Rothman of the University of Florida's Research Laboratory toward developing (and describing) a diver's lexicon. He listed the groups sampled and expressed the desire to obtain more samples from many more groups of divers, especially those engaged in operational tasks of salvage, repair, exploration and recovery. A systematized format for voice communication applicable to diving situations may provide the redundancy necessary for adequate speech information transmission until communication systems achieve distortion-free capabilities. Even under ideal transmission conditions the standardized format tends to help minimize semantic and linguistic confusion factors.

Dr. John Gill ventured a supposition that the previously mentioned techniques may still not solve all divers' speech problems. It may be that divers will not be chosen just as divers but a selection made from a population of those individuals who are highly-educated divers and who could learn to speak well. One parameter for selection is to determine if the diver has, or can achieve, a low vowel-consonant intensity ratio. Speakers who exhibit such characteristics tend to be more intelligible over most communication channels. This, of course, is the condition which engineers attempt to create artificially by the process of peak clipping.

Modification or Preprocessing of the Speech Signal

Additionally, Dr. Gill stated that when speech is to be processed by some "translator" device, the talker who employs "more-than-average" vocal force as he talks should result in enriching the overtone structure of his speech and allow the processing device to work on a signal having high "information" content, particularly if the loud speech is articulated in a very precise, "snappier", manner.

Another preprocessing method was postulated in which the diver would use whispered speech, which is not so closely related to volume velocity changes, and to amplify it strongly since the vowel-consonant intensity ratios are low. The dynamic range being relatively small may allow selective filtering or digital processing techniques to enable the whispered speech to be "translated" by simpler methods than time sampling. Consistantly produced whispered speech would require some training on the part of the diver, especially when he experiences episodes of extreme stress.

Other preprocessing techniques were mentioned including high-frequency preemphasis before translation, peak clipping and instantaneous speech-envelope compression. All of these modifications of the speech signal have been used in airborne voice communication systems with varying degrees of success. Their usefulness under helium-oxygen hyperbaric conditions needs experimental verification.

Communication Among Pressure Chamber Subjects

Subjects who are employed in hyperbaric chamber experimentation or indoctrination are reluctant to being restricted in their movements by earphone and microphone cords. There is some evidence provided by the analyses of Golden[22] and McLean[31] that divers tend to modify their speech over a period of time under hyperbaric conditions (at least to 205 feet equivalent depth). They found that over a period of two weeks, speakers, on the average, changed their speech in such a way that there was a shift downward in the second and third formants. The divers also report they were able to understand each other better as time progressed without resorting to microphones and translators (Sergeant[44] reports similar findings). This "adaptation" phenomenon needs investigation to determine what is being modified and if auditory feedback can speed the process by selectively eliminating from the feedback the factors which would allow the speaker to compensate in the "right" direction. However, some evidence as to the possibility of adequate subject intercommunications in hyperbaric chambers was given by Dr. John Gill concerning the British Admiralty's 1500-foot dive in which the subjects passed around a high quality microphone connected to a distortion-free amplification system which fed a loudspeaker in the chamber. He reports "good" communications, especially if the microphone-loudspeaker relationship was such as to cause no acoustic feedback. The newly developed gradient-type microphone described by Dr. Charles Morrow should be an aid in this situation.

Factors of Quick Solutions to Operational Problems

This portion of the discussion was initiated by several operational Navy representatives who asked what could be done quickly to provide speech communications for hard-hat divers. The need has long been apparent and they felt the situation is desperate. The principal disturbing factor seems to be a very high noise level inside the helmet, probably arising from the valving and venting of the breathing gases. Opinions were given that many things could be done, initially not requiring extensive research. Good consultation and application of engineering techniques should result in a useable communication system and improved communications environment. The prior mentioned gradient microphone should be tried as one component.

An extension of the above topic led to several expressions in which it was felt that there were communication gaps between Management, Operations and the Scientists. As seen by Operations, both Management and the Scientists tend to produce many fanciful solutions to immediate problems, liberally embellished with expressions of the difficulties in problem solution. Several representatives of Science, on the other hand, lamented the situation in which they felt both Management and Operations did not provide enough chances to do careful studies during operational dives. Whenever such chances did occur, there seemed to be insufficient time to obtain the requisite number of speech samples or to plan a satisfactory experiment. It was suggested by certain scientists and operations personnel that the most expedient and perhaps satisfactory solution to the communications gap was to establish personal contacts and to maintain and foster such contacts. The "personal approach" would not be an attempt to circumvent Management, since any studies or requests for aid must include Management, but some of the basic needs and feasibility statements could have preliminary planning, or possible sources of disagreement, smoothed.

Comments were made indicating that potential contractors and Navy personnel have been periodically confused because of an apparent lack of "singleness of purpose" or fractionation of efforts among various Navy organizations who have (and have had) an interest and/or responsibility concerning underwater communications. It was suggested a strong organization or central office be established which would coordinate, assign projects, and provide the funds for the best possible solutions to problems that the environment will allow. The above suggestions were made with the realization that money and priorities are an everchanging "balance-of-power" interaction, yet a single coordinator could minimize the struggle for program priorities among various Navy organizations who have, or have developed, an interest in underwater speech communications.

NavOP-23, the Supervisor of Salvage and Diving, the Navy Experimental Diving Unit, the Bureau of Medicine and Surgery, the Office of Naval Research, and others have been instrumental in initiating and maintaining efforts directed toward underwater communications. If delays in present efforts are encountered, or if the results of present efforts do not provide adequate systems, then perhaps a Task Unit should be created by NavOP-23, with members from appropriate Naval organizations, to remain in existence until divers can talk to each other easily and reliably.

At present, lacking such a central Task Unit, a suggested procedure for all Navy personnel involved in communication systems research and development is to maintain and foster strong interpersonal relationships with their management, hoping to influence them to increase the number of scientific-engineering discussions concerned with underwater communications. The results of the discussions should then be fed back to management as rapidly as possible. From such discussions, management, operations and science may be able to assign appropriate priority ranking to viable problem solutions with minimal delay which hopefully will accelerate operational capability.

In summary, the discussion in open forum seemed to indicate that the Navy should soon know whether it will have an interim all-diver speech communication system. If the systems now under test are not adequate, suggestions were made as to several "next steps", including: personnel selection and training; speech preprocessing methods; and alternative helium-speech translating methods. Suggestions were made by which information feedback to management, operations and science could be accelerated and utilized in planning and problem solution.

WORKSHOP SUMMARY AND COMMENTS

by W. Wathen-Dunn

My function, as I see it, is not to summarize in great detail what each speaker said, but rather to present an integrated picture that will tell us where we are now and show in what direction we ought to go to reach our objectives.

The problems we have are three in number. The first is the classical problem of describing and explaining the phenomena with which we have to deal. The second is that of trying to do something about the limitations they impose, and the third is to evaluate the results of our efforts.

In approaching these problems, I cannot emphasize too strongly the concept, advanced by several speakers, that it is vital to look at the over-all system, not just the components. It is also necessary to remember that the system includes human beings both as talkers and listeners and that this compounds our difficulties. A purely physical system exhibits much greater stability than one that includes people.

A diagram of the over-all system is given in Fig. 16. First, there is the speech generation process, and we need to know what the normal process is and the ways in which it is modified by its operating in a different pressure, having a different gas and being loaded differently. These are the effects of environment, which may also add noise and/or reverberation. The signal, thus modified and degraded, is converted to electrical form by a transducer which feeds a transmission system about which I shall not talk. At the receiving end, there is another

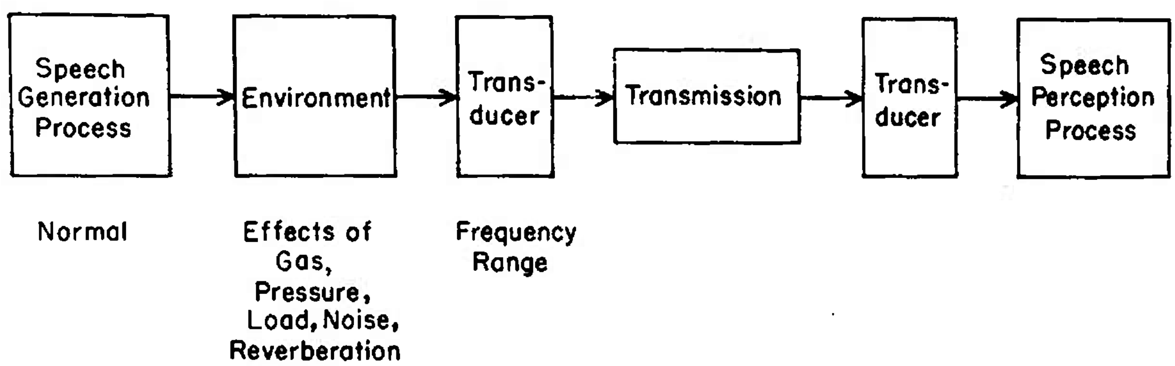


Fig. 16. (Wathen-Dunn) Diagram of "The Over-all System" for any voice communication.

transducer to convert the signal to audible sound and then, finally, the speech perception process.

We make a fundamental assumption, though no one here has stated it explicitly, that, if you can convert the short-term speech spectrum into some semblence of the form it would have had if the speaker had been talking in a normal environment, the result will be perceptually acceptable to the listener. On this assumption, we shall have to insert somewhere in the system, either before or after transmission, a device whose job it is to get the spectrum back into shape.

To understand what is required of this device, we must understand the speech generation process itself. Great insights have been gained by considering this to be a problem of network analysis and using methods developed by the electrical engineer to deal with it. He describes the process in terms of excitation functions that are applied to a "black box", as shown in Fig. 17. For our purposes, we need only recognize that speech requires a source of sound and that this is the excitation. There are two such sources: (1) voicing, produced by drawing the vocal cords together and forcing air to flow between them to make them vibrate; and/or (2) noise, produced by constricting the vocal tract at some point and forcing air through the constriction or, in some cases, closing the tract, building up air pressure behind the closure and then releasing it.

These excitations are time varying functions, and there is inevitably

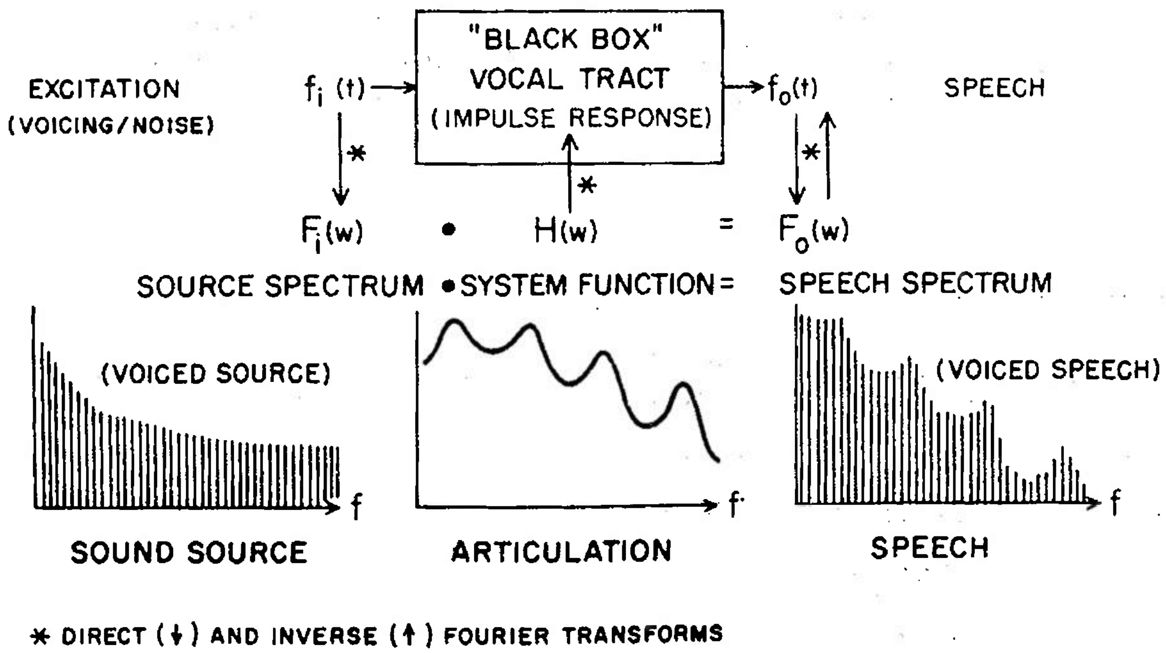


Fig. 17. (Wathen-Dunn) The concept of excitation functions and the "black box" for the speech generation process.

associated with each one a spectrum. Frequencies are generated in the audible range, and there are mathematical processes that can transform the time functions into functions of frequency. The simplest is the Fourier transform, and with it we can express all the frequencies in the excitation.

What are these frequencies? For voicing, the spectrum consists of a fundamental and a series of harmonics whose amplitudes fall off with frequency, as John Gill pointed out. The rate of fall-off is a function of vocal effort. It is more rapid for weak speech than for strong, and Gunnar Fant assumes it to be about 12 dB/octave for normal, conversational speech. The spectrum of the noise produced by turbulence in a constriction can be taken to be fairly uniform with frequency.

These spectra are applied to the vocal tract, which plays the role of the "black box." The vocal tract is a passive thing. It does not generate sound, but it can modify the sounds presented to it. Its transmission properties, characterized by a system function, attenuate some frequencies while reinforcing others and impose phase delays that likewise vary with frequency. In other words, the vocal tract affects the amplitude and phase of every frequency in the excitation. If we multiply the frequency function of the excitation by the system function of the vocal tract, we get the spectrum of the speech output, which may be converted to the output time waveform by an inverse Fourier transform. In doing this, radiation effects can be lumped with either the excitation spectrum or the system function.

That is the theory, but it might be most useful for us to think of the vocal tract as an acoustical transmission system that, like an organ pipe, has resonances. These resonances are called formants. They evince themselves as peaks, or humps, in the spectrum -- places where excitation frequencies are reinforced. The result for a typical vowel is shown in Fig. 17.

Lastly, not only is the vocal tract passive, but its system function changes continuously as we move from one articulatory configuration to another. Incidentally, our perception of speech includes the perception of articulation. We can mimic. If we hear a speech sound, we can manipulate our whole articulatory apparatus so as to repeat it, because knowledge of the articulatory position is implicit in our perception of the sound.

The foregoing describes the normal speech generation process. What modifications are caused by a different gas, pressure and loading? First, we are all aware that helium in the gas causes the formant frequencies to shift upward by reason of the increase in sound velocity, but it is important to note that this places these passive resonances in positions where there is less energy to excite them, because the excitation spectrum falls off with frequency. I assume it falls off in much the same way that it does for air, though I know of no one who can assure me on this point. Certainly, the fundamental frequency of vibration of the vocal cords is unchanged by the introduction of large amounts of helium at normal atmospheric pressure.

For high ambient pressures, we know that the minimum frequency of the first formant is raised so that there is no longer a linear relation between the value of any formant for air at sea level and its value for air at depth. Jan Lindqvist made the important point that the effect of this is reduced in helium speech, because in that case all the formant frequencies are raised. To the best of my knowledge, the effects of pressure on the excitation spectrum are unknown, though they ought to be, because significant changes in this spectrum, caused either by gas or pressure, could have important implications for how you restore speech to normalcy. Some sort of amplitude compensation as a function of frequency may be necessary.

You might also want to introduce another kind of amplitude compensation. With increased pressure, the amplitude of the voicing excitation increases as the square root of the pressure, whereas that of the noise remains unchanged.[9, 11] This makes voiced speech sounds louder than unvoiced ones, which might lead you to process the speech in one or the other of two different ways, depending on the voice-voiceless distinction.

A related question was raised in the discussion, namely, whether it is possible for a speaker to compensate by some learned modification in the way he talks. One suggestion would be to feed him sidetone from the output of the unscrambler, and from this he might learn to articulate in such a way as to improve the received intelligibility. This assumes that pressure has no appreciable effect on the motion of the articulators.

The external environment may consist of masks or helmets which, with their gas regulatory mechanisms, introduce several problems. A mask reacts with the vocal tract to distort the resonant system, and it may also impede the flow of exhaled gas so that voicing is affected. Gas supply and discharge mechanisms create noise that ought to be reduced as much as possible, for it further complicates the spectrum restoration process. Helmets have a serious resonance in an important part of the speech spectrum. A re-design of the interior shape of diving helmets and the introduction of some damping would be beneficial.

For helium speech it is necessary to have a wide-frequency-range system ahead of the unscrambler, but it seems that only recently has this been implemented. The microphone, of course, is the first component in such a system, and I gather that Dr. Morrow's microphone fills the bill. Provision of a response that rises with frequency, either in the microphone or in the associated pre-amplifier, can be used to compensate, at least in part, for the lower formant amplitudes in helium speech caused by the falling source spectrum.

Yesterday, Craig Allen indicated three different methods for unscrambling helium speech: (1) heterodyning; (2) "vocodering"; and (3) time-domain processing. The Navy Applied Sciences Lab unscrambler in the early 60's used the heterodyne method. The signal was broken up into two or three bands that were shifted downward by differing amounts to make them occupy roughly the proper spectral region. One difficulty with this method is that it does not necessarily preserve the harmonic structure of the voicing, and this is a source of distortion.

The Stockholm delegation gave a good résumé of vocoder techniques. A vocoder analyzes the input speech spectrum, dividing it into contiguous subbands with a set of bandpass filters, after which it takes a running measure of the acoustical energy in each band. At the receiver, it synthesizes a replica of the original speech by applying appropriate excitation to another set of contiguous bandpass filters, modulating the output of each one according to the running measure for that filter from the analyzer and summing the outputs. Ordinarily, the two sets of filters are identical, but the essential feature for our purposes is that the synthesizing filters can be made to occupy a lower and narrower frequency range and thus correct for helium distortions. The arrangement is inflexible, though, unless you can make the filters tunable in the way Mr. Allen suggested.

Most, if not all, of the unscramblers we listened to yesterday work in the time domain. A segment of normal voiced speech has a waveform like that

Fig. 18. (Wathen-Dunn) Acoustic waveforms of speech. a) Normal voice. b) Helium-speech.

shown in Fig. 18a. The principle peaks are caused by voice pulses, and the time between them is the fundamental period. The lesser vibrations in between result from the formants' ringing. A segment of voiced helium speech is illustrated in Fig. 18b. Here the formant vibrations occur much more rapidly and therefore crowd to the left, though the fundamental period remains unchanged. If you chop off the tail of the ringing at a suitable point and expand what is left to occupy the full voicing period, this stretches the waveform into some semblence of its original shape. This has to be done on a pitch synchronous basis, but the amount chopped off can be varied continuously, which provides flexibility.

Digital processing offers other means for dealing with helium speech. One is homomorphic processing, which includes cepstral processing. Recall (from Fig. 17) that the product $F_i(\omega) \cdot H(\omega) = F_O(\omega)$, the output speech spectrum. If we take the log of both sides of this equation, the left-hand side becomes an addition, $\log F_i(\omega) + \log H(\omega)$, which suggests the possibility that suitable processing of $\log F_O(\omega)$ could separate the excitation- and system-function components in some sense. This turns out to be partially true. If we transform $\log F_O(\omega)$ back into a time domain, t', we get what is called a "cepstrum." Concentrated about the origin is a function that represents the combination of source spectrum and vocal tract transfer characteristic. Further out is a series of peaks that are spaced from the origin and from each other by the fundamental period, and these are due to voicing. If we truncate this function so as to eliminate all the voicing peaks, the remaining portion adjacent to the origin can be transformed back into the frequency domain to give a smooth curve that represents the spectral effects of transfer characteristic and source spectrum.

Frank Quick worked with this curve.[37,38] He squeezed it to the left to give the formants their proper positions, exponentiated it to remove the effects of logging, and then dealt with the result as though it were simply a system function. He transformed this into the time domain to give an impulse response which he convolved with either regular or irregular pulses to reconstruct the speech. The output was far more intelligible than the original and might have been better if he had had a better tape of helium speech to start with.

A different method, based on predictive coding, has been developed by Bishnu Atal at Bell Labs. I shan't attempt to explain it, but it appears to yield a simple algorithm for getting the inverse filtering characteristic of the vocal tract, which is what we want to manipulate. Its simplicity allows it to be done in less computational time, and this gets to the heart of the matter. Digital processing is useful only if you can build dedicated computers that are fast enough to do the required processing in real time and that are small and cheap enough to make it practical. Computers that fulfill these requirements are not available at the moment, but the state-of-the-art seems to be progressing rapidly in that direction. Meanwhile, these processing methods are a powerful research tool.

Lastly, we come to evaluation. I think there are three aspects of the speech signal that are important to preserve in a communications system. The first, of course, is intelligibility, but a second is talker identity, and a third is what I call the emotional content of the speech. These convey what was said, who said it and how it was said. In helium speech we seem to be happy if we can preserve intelligibility, and we haven't worried about the other aspects. There are tests for evaluating a system for intelligibility, but it must be kept in mind that they do not yield really absolute results. The best you can do is rank-order several systems using the same crew and test conditions at the same time, and even then a 2% difference is not very meaningful. For testing unscramblers, Cdr. Bloom made the valid point that the talkers and listeners ought very likely to be divers.

There is just one more point. Closed message sets were discussed, and they might possibly increase intelligibility. However, from the point of view of information theory, it is the unexpected message -- the least probable one, the emergency one -- that carries the greatest amount of information, and this ought to be considered in constructing any closed message set.

REFERENCES

1. Barany, E., "Transposition of Speech Sounds," J. Acoust. Soc. Amer., 8, 217 (1937).

2. Beil, G., "Frequency Analysis of Vowels Produced in a Helium-Rich Atmosphere," J. Acoust. Soc. Amer., 34, 347-349 (1962).

3. Brubaker, R.S. and Wurst, J.W., "The Correction of Helium-Speech Distortion", H. R. B. Singer, Inc. Final Report, Science Park, State College, Penna., 1-35 (1967).

4. Copel, M., "Helium Voice Unscrambling", IEEE Trans. on Audio. and Electroacoust., 14, 122-126 (1966).

5. Delattre, P.C., Liberman, A.M. and Cooper, F.S., "Acoustic Loci and Transitional Cues for Consonants", J. Acoust. Soc. Amer., 27, 769-773 (1955).

6. Dublin, W.B., Baldes, E.J. and Williams, M.M.D., "Voice Changes with a Mixture of Helium and Oxygen", Proc. Staff Meeting, Mayo Clinic, 15, 586-588 (1940).

7. Dugan, J., Man Under the Sea, Collier Books, New York (1965).

8. Fairbanks, G., Everett, W.L. and Jaeger, R.D., "Method for Time or Frequency Compression of Speech", Trans. Inst. Radio Eng., AU2: 7-12 (1954).

9. Fant, G. and Lindqvist, J., "Pressure and Gas Mixture Effects on Divers' Speech", STL-Quarterly Progress and Status Report, Royal Institute of Technology (KTH), Stockholm, Sweden, 7-17 (1968).

10. Fant, G., Lindqvist, J., Sonneson, B. and Hollien, H., "Speech Distortion at High Pressures", Proc. of IV Internat. Cong. on Diver Physiol., Academic Press, New York, 293-299 (1971).

11. Fant, G. and Sonesson, B., "Speech at High Ambient Air Pressure", STL Quart. Prog. & Status Report, Royal Instit. of Technology (KTH), Stockholm, Sweden, 9-21 (1964).

12. Flower, R.A., "Final Report on Helium-Oxygen Voice Communication Study", GPL Division, General Precision, Inc., Pleasantville, New York (1965).

13. Flower, R. A., "Final Technical Report on Helium Speech Investigations", for Singer-Kearfott Div., Singer Co. to Office of Naval Research (ONR Project N00014-68-C-0330) (Dec 1969).

14. Flower, R.A., "Final Technical Report on Helium Voice Translator", Singer-Kearfott Div., Singer Co. (Internal Memorandum) (Jan 1971).

15. Fortune, R. G., "A Preliminary Study of Voice Communications while Breathing a Helium Mixture", USL Tech. Memo. 952-3-61, Naval Underwater Sound Laboratory, New London, Conn. (1961).

16. French, N.R. and Zinn, M.K., US Patent No.1, 761, 151 of 10 Dec 1924.

17. Fujimura, O. and Lindqvist, J., "Sweep-Tone Measurements of Vocal-Tract Characteristics", J. Acoust. Soc. Amer., 49, 541-558 (1971).

18. Gabor, D., J.I.E.E., Part III, 94, 369-387 (1947).

19. Gerstman, L.J., Gamertsfelder, G.R. and Roldberger, A., Breathing Mixture and Depth as Separate Effects on Helium Speech", J. Acoust. Soc. Amer., 40, 1284 (A) (1966).

20. Gill, J.S., Morris, R. and Edwards, M. G., "A Time-Domain Processor for Helium Speech", I.E.R.E. Conf. Proceedings, 19, 523-528 (1970).

21. Gill, J. S., Morris, R. and Edwards, M. G., British Patent Application 29400/70 (1970).

22. Golden, R. M., "Improving Naturalness and Intelligibility of Helium Speech using Vocoder Techniques", J. Acoust. Soc. Amer., 40, 621-624 (1966).

23. Hollien, H. and Coleman, R. F., "Diver-to-Diver Speech Intelligibility of Underwater Communication Systems", J. Acoust. Soc. Amer., 46, 115-116 (1969).

24. Hollien, H., Coleman, R.F. and Rothman, H., 'Evaluation of Diver Communication Systems by a Diver-to-Diver Technique", Inst. Electrical and Electronic Trans. on Communication Technique, 19, 403-409 (1971).

25. Hollien, H., Rothman, H.B. and Feinstein, S.H., "Annotated Bibliography on Underwater Communications", CSL/ONR Report No. 43, Communication Sciences Laboratory, Univ. of Florida, Gainesville, Fla. (1971).

26. Holywell, K. and Harvy, G., "Helium Speech", J. Acoust. Soc. Amer., 36, 210-211 (1964).

27. Integrated Electronics Corp., Pamphlet, "Diver Communication System with Helium Voice Processor", Huntington Station, New York (1968).

28. Integrated Electronics Corp., "Operation and Maintenance Manual for Diver Communication System Model 702A", P/N 5300 000, Huntington Station, New York (1969).

29. Lynch, G., "A Harmonic Analysis of Hydrogen Tones", Quart. J. Speech, 25, 57 (1939).

30. MacInnis, J.B., "Living Under the Sea", Scientific Amer., 214, 3 (1966).

31. McLean, D.J., "Analysis of Speech in a Helium-Oxygen Mixture Under Pressure", J. Acoust. Soc. Amer., 40, 625-627 (1966).

32. Morrow, C.T. and Brouns, A.J., "Speech Communication in Diving Masks. I. Acoustics of Microphones and Mask Cavities", J. Acoust. Soc. Amer., 50, 1-9 (1971).

33. Morrow, C.T. and Brouns, A.J., "Speech Communication in Diving Masks. II. Psychoacoustic and Supplementary Tests", J. Acoust. Soc. Amer., 50, 10-22 (1971).

34. Murry, T., "The Intelligibility of Processed Helium Speech at Depths Greater than 1000 Feet", NavSubMedRschLab Report 697, Naval Submarine Medical Center, Groton, Conn. (1971).

35. Murry, T., Sergeant, R.L. and Angermeier, C., "Navy Diver/Swimmer Vocabularies: Phonemic Intelligibility in Hyperbaric Environments", NavSubMedRschLab Report 648, Naval Submarine Medical Center, Groton, Conn. (1970).

36. Oppenheim, A.V., "Speech Spectrograms using the Fast Fourier Transform", IEEE Spectrum, 8, 57-62 (1970).

37. Quick, R.F., Jr., "Helium Speech Translation using Homomorphic Techniques", AFCRL Report 70-0424, Bedford, Mass. (1970).

38. Quick, R.F., Jr., "Application of Homomorphic Filtering to Helium Speech Translation", NEREM 70 Record, IEEE, Boston, Mass. 154-155 (1970).

39. Rich, E.L., "Progress Report: Development Survey of Helium Speech Unscrambling", Westinghouse Elect. Corp., Baltimore, Maryland (1966).

40. Roworth, D., Elect. Comm., 45, 18-23 (1970).

41. Sangster, F.L.J., Philips Tech. Review, 31, 97-110 (1970).

42. Scott, R.J., "Time Adjustment in Speech Synthesis", J. Acoust. Soc. Amer., 41, 60-65 (1967).

43. Sergeant, R.L., "Speech During Respiration of a Mixture of Helium and Oxygen", Aerospace Med., 34, 826-829 (1963).

44. Sergeant, R.L., "Speech Intelligibility during Prolonged Exposure to Helium-Oxygen", NavSubMedRschLab Memo Report 63-8, Naval Submarine Base, Groton, Conn. (1963).

45. Sergeant, R.L., "The Effect of Frequency Passband upon the Intelligibility of Helium-Speech in Noise", NavSubMedRschLab Report 480, Naval Submarine Medical Center, Groton, Conn. (1966).

46. Sergeant, R.L., "Phonemic Analysis of Consonants in Helium-Speech", J. Acoust. Soc. Amer., 41, 66-69 (1967).

47. Sergeant, R.L., "Limitations in Voice Communications during Deep Submergence-Helium Dives", Marine Sciences Instrumentation, 4, 405-411 (1968).

48. Sergeant, R.L., "CH 10: Distortion of Speech", in The Physiology and Medicine of Diving and

Compressed Air Work, Bennett, P.B. and Elliott, D.H. (Eds.), Bailliere, Tindall & Cassell Ltd., London, and Williams & Wilkins Co., Baltimore, 213-225 (1969).

49. Sinclair, T.F., "Man-in-the-Sea", Indust. Res., Ocean Systems, Inc., 74-79 (1965).

50. Stover, W.R., "Technique for Correcting Helium Speech Distortion", J. Acoust. Soc. Amer., 41, 70-74 (1967).

51. Westinghouse Elect. Corp., Sales Pamphlet for Helium Speech Unscrambler, Baltimore, Maryland (1968).

52. Wathen-Dunn, W., Michaels, S.B. and Lieberman, P., "Pitch of Speech from a Talker Supplied with Various Amounts of Helium", Paper J7, J. Acoust. Soc. Amer., 36, 2002 (A) (1964).

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) NAVAL SUBMARINE MEDICAL CENTER Naval Submarine Medical Research Laboratory	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

PROCESSING HELIUM SPEECH PROCEEDINGS OF A NAVY-SPONSORED WORKSHOP AUGUST 1971

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Interim report

5. AUTHOR(S) (First name, middle initial, last name)

Edited by
Russell L. SERGEANT, Ph. D. and LT Thomas Murry, MSC, USN

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
22 May 1972	68	52

8a. CONTRACT OR GRANT NO. b. PROJECT NO. BuMed -M4306.03-2020DAC5 c. ONR - RR 042-09-02 NR 197-003 d.	9a. ORIGINATOR'S REPORT NUMBER(S) NSMRL Report Number 708 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. DISTRIBUTION STATEMENT

Approved for public release; distribution unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Naval Submarine Medical Center, Box 600 Submarine Base Groton, Connecticut 06340

13. ABSTRACT

This report is a detailed summary of the proceedings of a workshop held during August 1971 on helium-speech processing. The meeting was sponsored jointly by the Office of Naval Research and the Bureau of Medicine and Surgery. It was held at the Naval Submarine Medical Research Laboratory in Groton, Connecticut. Approximately 40 participants assembled, including foreign scientists, U. S. Navy scientists and operational personnel, Navy and independent contractors, and speech scientists in the academic world, all who have been active in underwater communications. Formal papers were presented and discussed, a forum and discussion period was held, and a summary was presented. Progress and future developments toward reliable speech communication under hyperbaric helium-oxygen conditions were assessed. Concepts of helium-speech processing were advanced from the need of an unscrambler unit to one which includes the understanding and nature of the constraints allied to the unscrambler; that is, talker, listener, face mask and transducer. It was concluded that after a decade of research, the ability to correct hyperbaric helium speech finally exists. Now a system that is small, inexpensive, rugged and reliable must be designed and incorporated into diving operations.

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Helium speech						
Voice Communication						
Underwater communication						
Speech intelligibility						
Helium-speech unscramblers						
Hyperbaric speech						